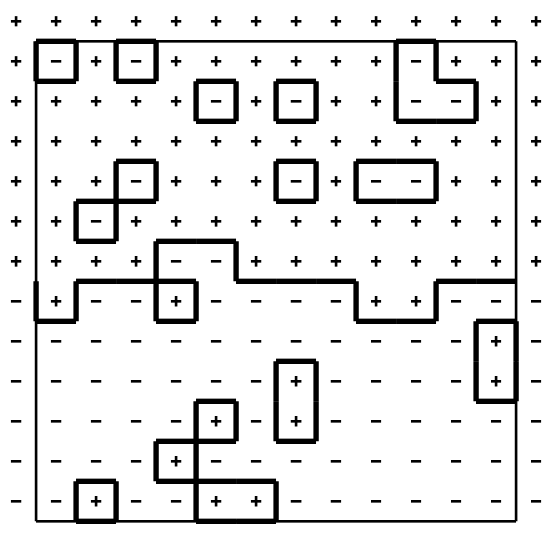
Giovanni Gallavotti

STATISTICAL MECHANICS

Short Treatise



Roma 1999

Short treatise of

Statistical Mechanics

Giovanni Gallavotti
Dipartimento di Fisica
Università di Roma *La Sapienza*
00185 Roma

Roma 1998

A Daniela, sempre

Preface

This book is the end result of a long story that started with my involvement as Coordinator of the Statistical Mechanics section of the Italian Encyclopedia of Physics.

An Italian edition collecting several papers that I wrote for the Encyclopedia appeared in September 1995, with the permission of the Encyclopedia and the sponsorship of Consiglio Nazionale delle Ricerche (CNR-GNFM).

The present work is not a translation of the Italian version but it overlaps with it: an important part of it (Ch.I,II,III,VIII) is still based on three articles written as entries for the —it Encicopledia della Fisica (namely: "*Meccanica Statistica*", "*Teoria degli Insiemi*" and "*Moto Browniano*") which make up about 29% of the present book and, furthermore, it still contains (with little editing and updating) my old review article on phase transitions (Ch.VI, published in La Rivista del Nuovo Cimento). In translating the ideas into English, I introduced many revisions and changes of perspective as well as new material (while also suppressing some other material).

The aim was to provide an analysis, intentionally as nontechnical as I was able to make it, of many fundamental questions of Statistical Mechanics, about two centuries after its birth. Only in a very few places have I entered into really technical details, mainly on subjects that I should know rather well or that I consider particularly important (the convergence of the Kirkwood-Salsburg equations, the existence of the thermodynamic limit, the exact soltution of the Ising model, and in part the exact solution of the six vertex models). The points of view expressed here were presented in innumerable lectures and talks mostly to my students in Roma during the last 25 years. They are not always "mainstream views"; but I am confident that they are not too far from the conventionally accepted "truth" and I do not consider it appropriate to list the differences from other treatments. I shall consider this book a success if it prompts comments (even if dictated by strong disagreement or dissatisfaction) on the (few) points that might be controversial. This would mean that the work has attained the goal of being noticed and of being worthy of criticism.

I hope that this work might be useful to students by bringing to their attention problems which, because of *"concreteness necessities"* (*i.e.* because such matters seem *useless*, or sometimes simply because of *lack of time*), are usually neglected even in graduate courses.

This does not mean that I intend to encourage students to look at questions dealing with the foundations of Physics. I rather believe that young students should *refrain* from such activities, which should, possibly, become a subject

of investigation after gaining an experience that only active and advanced research can provide (or at least the attempt at pursuing it over many years). And in any event I hope that the contents and the arguments I have selected will convey my appreciation for studies on the foundations that keep a strong character of concreteness. I hope, in fact, that this book will be considered concrete and far from speculative.

Not that students should not develop their own *philosophical* beliefs about the problems of the area of Physics that interests them. Although one should be aware that any philosophical belief on the foundations of Physics (and Science), no matter how clear and irrefutable it might appear to the person who developed it after long meditations and unending vigils, is very unlikely to look less than objectionable to any other person who is given a chance to think about it, it is nevertheless necessary, in order to grow original ideas or even to just perform work of good technical quality, to possess precise philosophical convictions on the *rerum natura*. Provided one is always willing to start afresh, avoiding, above all, thinking one has finally reached the *truth*, *unique, unchangeable and objective* (into whose existence only *vain* hope can be laid).

I am grateful to the *Enciclopedia Italiana* for having stimulated the beginning and the realization of this work, by assigning me the task of coordinating the Statistical Mechanics papers. I want to stress that the financial and cultural support from the *Enciclopedia* have been of invaluable aid. The atmosphere created by the Editors and by my colleagues in the few rooms of their facilities stimulated me deeply. It is important to remark on the rather unusual editorial enterprise they led to: it was not immediately animated by the logic of profit that moves the scientific book industry which is very concerned, at the same time, to avoid possible costly risks.

I want to thank G. Alippi, G. Altarelli, P. Dominici and V. Cappelletti who made a first version in Italian possible, mainly containing the Encyclopedia articles, by allowing the collection and reproduction of the texts of which the Encyclopedia retains the rights. I am indebted to V. Cappelletti for granting permission to include here the three entries I wrote for the *Enciclopedia delle Scienze Fisiche* (which is now published). I also thank the Nuovo Cimento for allowing the use of the 1972 review paper on the Ising model.

I am indebted for critical comments on the various drafts of the work, in particular, to G. Gentile whose comments have been an essential contribution to the revision of the manuscript; I am also indebted to several colleagues: P. Carta, E. Järvenpää, N. Nottingham and, furthermore, M. Campanino, V. Mastropietro, H. Spohn whose invaluable comments made the book more readable than it would otherwise have been.

Giovanni Gallavotti

Roma, January 1999

Index

Chapter I:

Classical Statistical Mechanics

.

§1.1. Introduction

Statistical mechanics poses the problem of deducing macroscopic properties of matter from the *atomic hypothesis*. According to the hypothesis matter consists of atoms or molecules that move subject to the laws of classical mechanics or of quantum mechanics.

Matter is therefore thought of as consisting of a very large number N of particles, essentially point masses, interacting via simple conservative forces.[1]

A *microscopic state* is described by specifying, at a given instant, the value of positions and momenta (or, equivalently, velocities) of *each* of the N particles. Hence one has to specify $3N + 3N$ coordinates that determine a point in *phase space*, in the sense of mechanics.

It does not seem that in the original viewpoint Boltzmann particles were really thought of as susceptible of assuming a $6N$ dimensional continuum of states, ([Bo74], p. 169):

Therefore if we wish to get a picture of the continuum in words, we first have to imagine a large, but finite number of particles with certain properties and investigate the behavior of the ensemble of such particles. Certain properties of the ensemble may approach a definite limit as we allow the number of particles ever more to increase and their size ever more to decrease. Of these properties one can then assert that they apply to a continuum, and in my opinion this is the only non-contradictory definition of a continuum with certain properties

and likewise the phase space itself is really thought of as divided into a finite number of very small *cells* of essentially equal dimensions, each of which determines the position and momentum of each particle with a *maximum precision*.

This should mean the maximum precision that the most perfect measurement apparatus can possibly provide. And a matter of principle arises: can we suppose that every lack of precision can be improved by improving the instruments we use?

If we believe this possibility then phase space cells, representing microscopic states with maximal precision, must be points and they must be conceived of as a $6N$ dimensional continuum. But since atoms and molecules are not directly observable one is legitimized in his doubts about being allowed to assume perfect measurability of momentum and position coordinates.

In fact in "recent" times the foundations of classical mechanics have been

---

[1] $N = 6.02 \times 10^{23}$ particles per mole = "Avogadro's number": this implies, for instance, that 1 cm^3 of Hydrogen, or of any other (perfect) gas, at normal conditions (1 *atm* at 0 °C) contains about 2.7×10^{19} molecules.

subject to intense critique and the *indetermination principle* postulates the theoretical impossibility of the simultaneous measurement of a component p of a particle momentum and of the corresponding component q of the position with respective precisions δp and δq without the constraint:

1.1.1 $$\delta p \delta q \ge h \tag{1.1.1}$$

where $h = 6.62 \times 10^{-27}\, erg \cdot sec$ is the *Planck's constant.*

Without attempting a discussion of the conceptual problems that the above brief and superficial comments raise it is better to proceed by imagining that the *microscopic states* of a N particles system are represented by phase space cells consisting in the points of R^{6N} with coordinates, (*e.g.* [Bo77]):

1.1.2 $$\begin{cases} p^\circ_\alpha - \delta p/2 \le p_\alpha \le p^\circ_\alpha + \delta p/2 \\ q^\circ_\alpha - \delta q/2 \le q_\alpha \le q^\circ_\alpha + \delta q/2 \end{cases} \qquad \alpha = 1, \ldots, 3N \tag{1.1.2}$$

if p_1, p_2, p_3 are the momentum coordinates of the first particle, p_4, p_5, p_6 of the second, *etc*, and q_1, q_2, q_3 are the position coordinates of the first particle, q_4, q_5, q_6 of the second, *etc*... The coordinate p°_α and q°_α are used to identify the center of the cell, hence the cell itself.

The cell size will be supposed to be such that:

1.1.3 $$\delta p \delta q = h \tag{1.1.3}$$

where h is an *a priori* arbitrary constant, which it is convenient *not to fix* because it is interesting (for the reasons just given) to see how the theory depends upon it. Here the meaning of h is that of a limitation to the precision that is assumed to be possible when measuring a pair of corresponding position and momentum coordinates.

Therefore the space of the microscopic states is the collection of the cubic cells Δ, with volume h^{3N} into which we imagine that the phase space is divided. *By assumption* it has no meaning to pose the problem of attempting to determine the microscopic state with a greater precision.

The *optimistic viewpoint* of orthodox statistical mechanics (which admits perfect simultaneous measurements of positions and momenta as possible) will be obtained by considering, in the more general theory with $h > 0$, the limit as $h \to 0$, *which will mean* $\delta p = \lambda p_0, \delta q = \lambda q_0$, *with* p_0, q_0 *fixed and* $\lambda \to 0$.

Even if we wish to ignore (one should not!) the development of quantum mechanics, the real possibility of the situation in which $h = 0$ cannot be directly checked because of the practical impossibility of observing an individual atom with infinite precision (or just with "great" precision).

§1.2. Microscopic Dynamics

The atomic hypothesis, apart from supposing the existence of atoms and molecules, assumes also that their motions are governed by a deterministic *law of motion.*

This hypothesis can be imposed by thinking that there is a map S:

1.2.1
$$S\Delta = \Delta' \tag{1.2.1}$$

transforming the phase space cells into each other and describing the system dynamics.

If at time t the state of the system is microscopically determined by the phase space cell Δ, then at a later time $t+\tau$ it will be determined by the cell Δ'. Here τ is a time step extremely small compared to the macroscopic time intervals over which the system evolution is followed by an observer: it is, nevertheless, a time interval directly accessible to observation, at least in principle.

The evolution law S is not arbitrary: it must satisfy some fundamental properties; namely it must agree with the laws of mechanics in order to properly enact the deterministic principle which is basic to the atomic hypothesis.

This means, in essence, that one can associate with each phase space cell three fundamental dynamical quantities: the *kinetic energy*, the *potential energy* and the *total energy*, respectively denoted by $K(\Delta), \Phi(\Delta), E(\Delta)$.

For simplicity assume the system to consist of N identical particles with mass m, pairwise interacting via a conservative force with potential energy φ. If Δ is the phase space cell determined by (see (1.1.2)) $(\underline{p}^\circ, \underline{q}^\circ)$, then the above basic quantities are defined respectively by:

1.2.2
$$\begin{aligned} K(\underline{p}^\circ) \equiv K(\Delta) &= \sum_{i=1}^{N} (\underline{p}_i^\circ)^2/2m \qquad && \underline{p}_i^\circ = (p_{3i-2}^\circ, p_{3i-1}^\circ, p_{3i}^\circ) \\ \Phi(\underline{q}^\circ) \equiv \Phi(\Delta) &= \sum_{i<j}^{1,N} \varphi(\underline{q}_i^\circ - \underline{q}_j^\circ) \qquad && \underline{q}_i^\circ = (q_{3i-2}^\circ, q_{3i-1}^\circ, q_{3i}^\circ) \\ E(\underline{p}^\circ, \underline{q}^\circ) \equiv E(\Delta) &= K(\underline{p}^\circ) + \Phi(\underline{q}^\circ) \end{aligned} \tag{1.2.2}$$

where $\underline{p}_i^\circ = (\underline{p}_{3i-2}^\circ, p_{3i-1}^\circ, p_{3i}^\circ)$, $\underline{q}_i^\circ = (\underline{q}_{3i-2}^\circ, q_{3i-1}^\circ, q_{3i}^\circ)$ are the momentum and position of the i-th particle, $i = 1, 2, ...N$, in the *microscopic state* corresponding to the center $(\underline{p}^\circ, \underline{q}^\circ)$ of Δ.

Replacing $\underline{p}^\circ, \underline{q}^\circ$, *i.e.* the center of Δ, by another point $(\underline{p}, \underline{q})$ in Δ one obtains values $K(\underline{p}), \Phi(\underline{q}), E(\underline{p}, \underline{q})$ for the kinetic, potential and total energies different from $K(\Delta)$,$\Phi(\Delta)$,$E(\Delta)$; *however* such a difference has to be *non observable*: otherwise the cells Δ would not be the smallest ones to be observable, as supposed above.

If τ is a fixed time interval and we consider the solutions of Hamilton's equations of motion:

1.2.3
$$\dot{\underline{q}} = \frac{\partial E}{\partial \underline{p}}(\underline{p}, \underline{q}), \qquad \dot{\underline{p}} = -\frac{\partial E}{\partial \underline{q}}(\underline{p}, \underline{q}) \tag{1.2.3}$$

with initial data $(\underline{p}^\circ, \underline{q}^\circ)$ at time 0 the point $(\underline{p}^\circ, \underline{q}^\circ)$ will evolve in time τ into a point $S_\tau(p^\circ, q^\circ) = (\underline{p}', \underline{q}') = (S_\tau(\underline{p}^\circ, \underline{q}^\circ))$. One then defines S so that

$S\Delta = \Delta'$ if Δ' is the cell containing $(\underline{p}', \underline{q}')$. The evolution (1.2.3) may send a few particles outside the volume V, cubic for simplicity, that we imagine to contain the particles: one has therefore to supplement (1.2.3) by *boundary conditions* that will tell us the physical nature of the walls of V.

They could be *reflecting*, if the collisions with the walls are elastic, or *periodic* if the opposite faces of the region V are identified (a very convenient "mathematical fiction", useful to test various models and to minimize the "finite size" effects, *i.e.* dependence of observations on system size).

One cannot, however, escape some questions of principle on the structure of the map S that it is convenient not to ignore, although their deep understanding may become a necessity only on a second reading.

First we shall neglect the possibility that $(\underline{p}', \underline{q}')$ is on the boundary of a cell (a case in which Δ' is not uniquely determined, but which can be avoided by imagining that the cells walls are slightly deformed).

More important, in fact crucial, is the question of whether $S\Delta_1 = S\Delta_2$ implies $\Delta_1 = \Delta_2$: the latter is a property which is certainly true in the case of point cells ($h = 0$), because of the uniqueness of the solutions of differential equations. It has an obvious intuitive meaning and an interest due to its relation with reversibility of motion.

In the following analysis a key role is played by Liouville's theorem which tells us that the transformation mapping a generic initial datum $(\underline{p}, \underline{q})$ into the configuration $(\underline{p}', \underline{q}') = S(\underline{p}, \underline{q})$ is a *volume preserving* transformation.

This means that the set of initial data $(\underline{p}, \underline{q})$ in Δ evolves in the time τ into a set $\tilde{\Delta}$ with volume equal to that of Δ. Although having the same volume of Δ *it will no longer have the same form* of a square parallelepiped with dimensions δp or δq. For h small it will be a rather small parallelepiped obtained from Δ via a linear transformation that expands in certain directions while contracting in others.

It is also clear that in order that the representation of the microscopic states of the system be consistent it is necessary to impose some non trivial conditions on the time interval, so far unspecified, that elapses between successive (thought) observations of the motions. Such conditions can be understood via the following reasoning.

Suppose that h is very small (actually by this we mean, here and below, that *both* δp and δq are small) so that the region $\tilde{\Delta}$ can be regarded as obtained by translating Δ and possibly by deforming it via a linear dilatation in some directions or a linear contraction in others (contraction and dilatation *balance* each other because, as remarked, the volume remains constant). This is easily realized if h is small enough since the solutions to ordinary differential equations can always be thought of, locally, as linear transformations close to the identity, for small evolution times τ. Then:

i) If S dilates and contracts in various directions, even by a small amount, there must necessarily exist pairs of distinct cells $\Delta_1 \neq \Delta_2$ for which $S\Delta_1 = S\Delta_2$: an example is provided by the map of the plane transforming (x, y) into $S(x, y) = ((1+\varepsilon)^{-1}x, (1+\varepsilon)y)$, $\varepsilon > 0$ and its action on the lattice of

the integers. Assuming that one decides that the cell Δ' into which a given cell Δ evolves is, among those which intersect its image $S\Delta$, the one which had the largest intersection with it, then indetermination arises for a set of cells spaced by about ε^{-1}.

It is therefore necessary that τ be small, say:

1.2.4 $$\tau < \vartheta_+ \qquad (1.2.4)$$

with ϑ_+ such that the map S (associated with S_τ, see (1.2.2), hence close to the identity) produces contractions and expansions of Δ that can be neglected for the *large majority* of the cells Δ. Only in this way will it be possible that $S\Delta_1 = S\Delta_2$ with $\Delta_1 \neq \Delta_2$ for just a small fraction of the cells and, hence, one can hope that this possibility is negligible.

It should be remarked explicitly that the above point of view is systematically taken by physicists performing numerical experiments. *Phase space is represented in computers as a finite*, but very large, set of points whose positions are changed by the time evolution (how many depends on the precision of the representation of the reals). Even if the system studied is modeled by a nice differential equation with global uniqueness and existence of solutions, the computer program, while trying to generate a permutation of phase space points, will commit errors, *i.e.* two distinct points will be sent to the same point (we do not talk here of round-off errors, which are not really errors as they are *a priori* known, in principle): one thus hopes that such errors are rare enough to be negligible. This seems inevitable except in some remarkable cases, the only nontrivial one I know of being in [LV93].

ii) *But* τ cannot be too small either, if one wishes to maintain coherently the point of view that microscopic states are described by phase space cells. In fact to a cell Δ is associated a natural time scale $\vartheta_-(\Delta)$: which can be defined as the minimum time in order that Δ becomes distinguishable from the cell into which it evolves in time $\vartheta_-(\Delta)$. And τ must be necessarily larger than the latter *minimum* time scale:

1.2.5 $$\vartheta_-(\Delta) < \tau \qquad (1.2.5)$$

(otherwise we have Zeno's paradox and nothing moves).

Summarizing we can say that in order to be able to define the dynamic evolution as a map permuting the phase space cells it must be that τ be chosen so that:

1.2.6 $$\vartheta_- = {}''\max_{\Delta}{}''\vartheta_-(\Delta) \le \tau \le \vartheta_+ \qquad (1.2.6)$$

where the quotes mean that the maximum has to be taken as Δ varies within the "majority" of the cells, where one can suppose that $\Delta_1 \neq \Delta_2$ implies $S\Delta_1 \neq S\Delta_2$.

One should realize that if φ is a "reasonable" molecular potential (a typical model for φ is, for instance, the *Lennard-Jones potential* with *intensity* ε and *range* r_0 given by: $\varphi(r) = 4\varepsilon((\frac{r_0}{r})^{12} - (\frac{r_0}{r})^6)$) it will generically be that:

1.2.7
$$\lim_{h\to 0} \vartheta_- = 0 \tag{1.2.7}$$

while for small h the right hand side of (1.2.6) (which has a purely kinematical nature) becomes h independent.

Hence it will be possible, at least in the limit in which h tends to 0, to define a τ so that (1.2.4), (1.2.5) hold; *i.e.* it will be possible to fulfill the above consistency criteria for the describing microscopic states of the system via finite cells.

On the other hand if $h > 0$, and *a posteriori* one should think that $h = 6.62 \times 10^{-27}$ *erg sec*, the question we are discussing becomes quite delicate: were it not because we do not know what we should understand when we say *the "large majority"* of the phase space cells.

In fact, on the basis of the results of the theory it will become possible to evaluate the influence on the results themselves of the existence of pairs of cells $\Delta_1 \neq \Delta_2$ with $S\Delta_1 = S\Delta_2$.

Logically at this point the analysis of the question should be postponed until the consequences of the hypotheses that we are assuming allow us to reexamine it. It is nevertheless useful, in order to better grasp the delicate nature of the problem and the orders of magnitude involved, to anticipate some of the basic results and to provide estimates of $\vartheta_\pm$': readers preferring to think in purely classical terms, by imagining that $h = 0$ on the basis of a dogmatic interpretation of the (classical) atomic hypothesis, can skip the discussion and proceed by systematically taking the limit as $h \to 0$ of the theory that follows.

It is however worth stressing that setting $h = 0$ is an illusory simplification avoiding posing a problem that is today well known to be deep. Assuming that, at least in principle, it should be possible to measure exactly positions and momenta of a very large number of molecules (or even of a single one) means supposing it is possible to perform a physical operation that *no one* would be able to perform. It was the obvious difficulty, one should recall, of such an operation that in the last century made it hard for some to accept the atomic hypothesis.

Coming to the problem of providing an idea of the orders of magnitude of $\vartheta_\pm$ one can interpret "max" in (1.2.6) as evaluated by considering as *typical* cells those for which the momenta and the reciprocal distances of the particles take values "close" to their "average values". The theory of *statistical ensembles* (see below) will lead to a natural probability distribution giving the probability of each cell in phase space, when the system is in macroscopic equilibrium. Therefore we shall be able to compute, by using this probability distribution, the average values of various quantities in terms of macroscopic quantities like the absolute temperature T, the particle mass m, the particle number N, and the volume V available to the system.

The main property of the probability distributions of the microscopic states observed in a situation in which the macroscopic state of the system is in equilibrium is that the average velocity and average momentum $\overline{v}, \overline{p}$ will be related to the temperature by:

1.2.8

$$\overline{p} = m\overline{v} = \sqrt{3mk_BT}, \qquad m\overline{v}^2 = 3k_BT \tag{1.2.8}$$

where $k_B = 1.38 \times 10^{-16}\, erg\, {}^\circ K^{-1}$ is the universal Boltzmann's constant.

Other relevant quantities are the characteristic parameters of the interaction, *i.e.* the *strength* ε with the dimension of energy and the *range* r_0 with the dimension of length. It follows from the developments of the theory of equilibrium statistical mechanics, *independently of the particular form of* $\varphi(r)$ (as long as it is "reasonable", like for instance the above mentioned Lennard-Jones potential), that $\varepsilon \cong k_B T_c^0$ where T_c^0 is the critical liquefaction temperature and r_0 is of the order of the molecular diameter (between $2 \times 10^{-8}\, cm$ and $4 \times 10^{-8}\, cm$ in the simplest gases like H_2, He, O_2, CO_2, see Chap.V).

We estimate ϑ_+ first (the time scale over which expansion and contraction of a phase space cell become sensible) looking at a typical cell where one can assume that the particles evolve in time without undergoing multiple collisions. In such a situation the relative variation of a linear dimension of Δ in the time τ will be, for small τ, proportional to τ and it may depend on $\varepsilon, m, r_0, \overline{v}$: the pure numbers related to τ and to the phase space dilatations (*i.e.* to the derivatives of the forces appearing in the equations of motion) that one can form with the above quantities are $\tau(\frac{\varepsilon}{mr_0^2})^{1/2}$ and $\tau(\frac{m\,\overline{v}^2}{mr_0^2})^{1/2}$. Hence the phase space changes in volume will be negligible, recalling that $m\,\overline{v}^2 = 3k_BT$, from (1.2.8), and setting $\varepsilon \equiv k_B T_c^0$, provided:

1.2.9

$$\tau < \min\Big(\Big(\frac{mr_0^2}{k_B T_c^0}\Big)^{1/2}, \Big(\frac{mr_0^2}{k_B T}\Big)^{1/2} \Big) \equiv \vartheta_+. \tag{1.2.9}$$

The condition $\tau\sqrt{\varepsilon/mr_0^2} < 1$ means that, even during a microscopic collision taking place while the time τ elapses, there is no sensible expansion while the second condition $\tau < r_0/\,\overline{v}$ means that the time τ is short with respect to the total collision duration (which therefore takes several units of τ to be completed).

To estimate ϑ_- (the time scale over which a cell evolves enough to be distinguishable from itself) note that, given Δ, the coordinates p_α, q_α of the phase space points in the cell Δ change obeying the Hamiltonian equations of motion, in the time τ, by:

1.2.10

$$\begin{aligned} |\delta q_\alpha| &\cong |\tau \frac{\partial E}{\partial p_\alpha}(\underline{p}, \underline{q})| = \tau \frac{\delta_\alpha^{(1)} E}{\delta p} \\ |\delta p_\alpha| &\cong |-\tau \frac{\partial E}{\partial q_\alpha}(\underline{p}, \underline{q})| = \tau \frac{\delta_\alpha^{(2)} E}{\delta q} \end{aligned} \tag{1.2.10}$$

where $\delta^{(1)}_\alpha E$, $\delta^{(2)}_\alpha E$ are the variations of the energy E in the cell Δ when the coordinates p_α or, respectively, q_α vary by the amount δp or δq, *i.e.* they vary by a quantity equaling the linear dimensions of the cell Δ, while the others stay constant (so that the variations in (1.2.10) are related to the partial derivatives of the energy function E).

Defining therefore the *energy indetermination*, which we denote by $\delta E(\Delta)$, in the cell Δ as:

1.2.11 $$\delta E(\Delta) = \max_\alpha (\delta^{(1)}_\alpha E, \delta^{(2)}_\alpha E) \tag{1.2.11}$$

we see that the minimum time $\vartheta_-(\Delta)$ that one has to wait in order to see that the cell evolves into a cell which is *distinguishable* from Δ itself is:

1.2.12 $$\vartheta_-(\Delta) \max_\alpha \frac{\delta^{(1)}_\alpha E}{\delta p} \ge \delta q, \quad \text{or} \quad \vartheta_-(\Delta) \max_\alpha \frac{\delta^{(2)}_\alpha E}{\delta q} \ge \delta p \tag{1.2.12}$$

in fact, δp and δq being the linear dimensions of Δ, Eq. (1.2.12) just says that at least one of the sides of Δ has moved away by a quantity of the order of its length (thus becoming distinguishable from itself).

Since we set $\delta p\, \delta q = h$ one deduces from (1.2.11),(1.2.12):

1.2.13 $$\vartheta_-(\Delta)\delta E(\Delta) \ge h \tag{1.2.13}$$

and $\vartheta_- \equiv \delta t$ can be chosen so that, introducing the notation:

1.2.14 $$\delta E = {}''\min''_\Delta \, \delta E(\Delta) \tag{1.2.14}$$

we have:

1.2.15 $$\delta t \delta E = h\,. \tag{1.2.15}$$

We can therefore see, on the basis of (1.2.9),(1.2.15), whether or not an interval $(\vartheta_-, \vartheta_+)$ admissible for τ exists. We can in fact imagine that $\delta p \simeq \overline{p}$, hence $\delta E = \overline{p}\, \delta p/m \cong \bar{p}^2/m = 3k_BT$ and:

1.2.16 $$\vartheta_- \equiv h/k_BT\,. \tag{1.2.16}$$

Equation (1.2.16) gives a remarkable interpretation of the time scale h/k_BT: it is the time necessary so that a phase space cell, typical among those describing the microscopic equilibrium states at temperature T, becomes distinguishable from itself.

One can say, differently, that ϑ_- is determined by the size of $\dot{p}, \dot{q}$, *i.e.* by the size of the first derivatives of the Hamiltonian, while ϑ_+ is related to the phase space expansion, *i.e.* to the second derivatives of the Hamiltonian.

With some algebra one derives, from (1.2.9), (1.2.16):

1.2.17 $$\vartheta_+/\vartheta_- = (mr_0^2 k_B T_c^0/h^2)^{1/2} \, \min(T/T_c^0, (T/T_c^0)^{1/2}). \tag{1.2.17}$$

Therefore it is clear that the relation $\vartheta_+/\vartheta_- > 1$, *necessary for a consistent description* of the microscopic states in terms of phase space cells, will be

satisfied for large T, say $T \gg T_0$, *but not for* small T (unless one takes $h = 0$). And from the expression just derived for the ratio ϑ_+/ϑ_- one gets $\vartheta_+/\vartheta_- > 1$ if $T > T_0$ with:

1.2.18
$$T_0/T_c^0 = \max\Big(\frac{h}{(mr_0^2 k_B T_c^0)^{1/2}}\,,\ \frac{h^2}{((mr_0^2 k_B T_c^0)^{1/2})^2}\Big)\,. \tag{1.2.18}$$

Table 1.1 below gives an idea of the orders of magnitude: it is elaborated having chosen $h = 6.62 \times 10^{-27}\, erg\, sec$, *i.e.* Planck's constant. A similar table would be derived if, *ignoring Planck's constant suggested as a natural action unit from quantum mechanics, we took* $\delta p \sim \overline{p}$, $\delta q \sim \sqrt[3]{V/N}$, as Boltzmann himself did when performing various conceptual calculations, [Bo96], [Bo97], in his attempts to explain why his physical theory did not contradict mathematical logic.

In fact with the latter choice the action unit $\delta p\, \delta q \cong \overline{p}\sqrt[3]{V/N}$ would be, in "reasonable cases" ($1\, cm^3$ of hydrogen, $m = 3.34 \times 10^{-24} g$, $T = 273\,^\circ$K, $N = 2.7 \times 10^{19}$, $k_B = 1.38 \times 10^{-16}\, erg^o K^{-1}$), of the *same order of magnitude* as Planck's constant, namely it would be $\delta p\, \delta q \cong 2.04 \times 10^{-25}\, erg \cdot sec.$

The corresponding order of magnitude of ϑ_- is $\vartheta_- \cong 5.4\ 10^{-12}$ sec.

The sizes of the estimates for T_0/T_c^0 in the table show that the question of logical consistency of the microscopic states representation in terms of phase space cells permuted by the dynamics, if taken literally, depends in a very sensitive way on the value of h and, in any event, it is *doomed to inconsistency* if $T \to 0$ and $\varepsilon \neq 0$ (hence $\vartheta_- \to +\infty$ and $\vartheta_+ \to \sqrt{mr_0^2/\varepsilon} < +\infty$).

Table 1.1: Orders of magnitude (N_A denotes Avogadro's number)

	r_0	$\varepsilon = k_B T_c^0$	T_0/T_c^0	$T_c^{\rm true}$	A	B	m
	10^{-8}cm	10^{-14}erg		°K	10^{-11} cgs	cgs	1.17 10^{-24}g
H_2	2.76	0.57	4.3	33.2	2.46	26.7	2
CO_2	3.23	5.25	0.12	304.1	36.	42.7	44
He	2.65	0.08	15.	5.19	0.33	23.5	4
N_2	3.12	2.17	0.25	126.	13.4	38.6	28
O_2	2.93	2.65	0.23	154.3	13.6	31.9	32

The columns A, B give empirical data, directly accessible from experiments and expressed in *cgs* units (*i.e.* A in $erg \cdot cm^3$ and B in cm^3), of the van der Waals' equation of state. If $n = N/N_A$ = number of moles, $R = k_B N_A$, see §5.1 for $(*), (**)$ below, then the equation of state is:

$$(P + An^2/V^2)(V - nB) = nRT \tag{$*$}$$

which is supposed here in order to derive values for ε, r_0 via the relations:

$$(B/N_A) = 4\frac{4\pi}{3}\Big(\frac{r_0}{2}\Big)^3 = 4v_0 \qquad A/N_A^2 = \frac{32}{3}\varepsilon v_0 \tag{$**$}$$

which lead to the expressions (see §4.3) $r_0 = (3B/2\pi N_A)^{1/3}, \varepsilon = 3A/8BN_A = 81\, k_B T_c^0/64$; T_c^{true} = experimental value of the critical temperature $\sim T_c^0$.

In this book I will choose the attitude of *not* attempting to discuss which would be the structure of a statistical mechanics theory of phenomena below T_0 if a strict, "axiomatic", classical viewpoint was taken assuming $\delta p = 0, \delta q = 0$: the theory would be extremely complicated as discovered in the famous simulation [FPU55] and it is still not well understood even though it is full of very interesting phenomena, see [GS72], [Be94], [Be97] and Chap.III, §3.2.

§1.3. Time Averages and the Ergodic Hypothesis

We are led, therefore, to describe a mechanical system of N identical mass m particles (at least at not too low temperatures, $T > T_0$, see (1.2.18)) in terms of (a) an *energy function* ("Hamiltonian") defined on the $6N$-dimensional phase space and (b) a subdivision of such a space into cells Δ of equal volume h^{3N}, whose size is related to the highest precision with which we presume to be able to measure positions and momenta or times and energies.

Time evolution is studied on time intervals multiples of a unit τ: large compared to the time scale δt associated with the cell decomposition of phase space by (1.2.15), (1.2.16) and small compared with the *collision* time scale (1.2.9): see [Bo74], p. 44, 227. In this situation time evolution can be regarded as a permutation of the cells with given energy: we neglect, in fact, on the basis of the analysis in §1.2 the possibility that there may be a small fraction of different cells evolving into the same cell.

In this context we ask what will be the qualitative behavior of the system with an energy "fixed" macroscopically, *i.e.* in an interval between $E - DE$ and E, if its observations are timed at intervals τ and the quantity DE is macroscopically small but $DE \gg \delta E = h/\delta t$; see (1.2.15), (1.2.16).

Boltzmann assumed, very boldly, that *in the interesting cases* the *ergodic hypothesis* held, according to which ([Bo71], [Bo84], [Ma79]):

Ergodic hypothesis: the action of the evolution transformation S, as a cell permutation of the phase space cells on the surface of constant energy, is a one cycle permutation of the $\mathcal{N}$ phase space cells with the given energy:

1.3.1 $$S\Delta_k = \Delta_{k+1} \qquad k = 1, 2, \ldots, \mathcal{N} \tag{1.3.1}$$

if the cells are suitably enumerated (and $\Delta_{\mathcal{N}+1} \equiv \Delta_1$).

In other words as time evolves every cell evolves, visiting successively all other cells with equal energy. The action of S is the simplest thinkable permutation!

Even if not strictly true this should hold *at least for the purpose of computing the time averages of the observables relevant for the macroscopic properties of the system.*

The basis for such a celebrated (and much criticized) hypothesis rests on its conceptual simplicity: it says that in the system under analysis all cells with the same energy are equivalent.

There are cases (already well known to Boltzmann, [Bo84]) in which the hypothesis is manifestly false: for instance if the system is enclosed in a perfect spherical container then the evolution keeps the angular momentum $\underline{M}(\Delta) = \sum_{i=1}^{N} \underline{q}_i^\circ \wedge \underline{p}_i^\circ$ constant. Hence cells with a different total angular momentum cannot evolve into each other.

This extends to the cases in which the system admits other constants of motion, besides the energy, because the evolving cells must keep the constants of motion equal to their initial values. And this means that the existence of other constants of motion besides the energy is, essentially, the most general case in which the ergodic hypothesis fails: in fact when the evolution is not a single cycle permutation of the phase space cells with given energy, then one can decompose it into cycles. One can correspondingly define a function A by associating with each cell of the same cycle the very same (arbitrarily chosen) value of A, different from that of cells of any other cycle.

Obviously the function A so defined is a constant of motion that can play the same role as the angular momentum in the previous example.

Thus, if the ergodic hypothesis failed to be verified, then the system would be subject to other conservation laws, besides that of the energy. In such cases it would be natural to imagine that *all* the conserved quantities were fixed and to ask oneself which are the qualitative properties of the motions with energy E, when all the other constants of motion are *also* fixed. Clearly in this situation the motion will be *by construction* a simple cyclic permutation of all the cells compatible with the prefixed energy and other constants of motion values.

Hence it is convenient to define formally the notion of *ergodic* probability distribution on phase space:

Definition: a set of phase space cells is ergodic if S maps it into itself and if S acting on the set of cells is a one-cycle permutation of them.

Therefore, in some sense, *the ergodic hypothesis would not be restrictive* and it would simply become the statement that one studies the motion after having *a priori* fixed all the values of the constants of motion.

The latter remark, as Boltzmann himself realized, does not make less interesting the concrete question of determining whether a system is ergodic in the strict sense of the ergodic hypothesis (*i.e.* no other constants of motion besides the energy). On the contrary it serves well to put in evidence some subtle and deep aspects of the problem.

In fact the decomposition of S into cycles (*ergodic decomposition of S*) might turn out to be so involved and intricated to render its construction practically impossible, *i.e.* useless for practical purposes. This would happen if the regions of phase space corresponding to the various cycles were (at least in some directions) of microscopic size or of size much smaller than a macroscopic size, or if they were very irregular on a microscopic scale: a quite different a situation if compared to the above simple example of the conservation of angular momentum.

It is not at all inconceivable that in interesting systems there could be very complicated constants of motion, without a direct macroscopic physical meaning: important examples are discussed in [Za89].

Therefore the *ergodic problem*, *i.e.* the problem of verifying the validity of the ergodic hypothesis for specific systems, in cases in which no particular symmetry properties can be invoked to imply the existence of other constants of motion, is a problem that remains to be understood on a case-by-case analysis. A satisfactory solution would be the proof of strict validity of the ergodic hypothesis or the possibility of identifying the cycles of S via level surfaces of simple functions admitting a macroscopic physical meaning (*e.g.* simple constants of motion associated with macroscopic "conservation laws", as in the case of the angular momentum illustrated above).

It is useful to stress that one should not think that there are no other simple and interesting cases in which the ergodic hypothesis is manifestly false. The most classical example is the *chain of harmonic oscillators*: described by:

1.3.2
$$T = \sum_{i=1}^{N} p_i^2/2m, \qquad \Phi = \sum_{i=1}^{N} m(q_{i+1} - q_i)^2/2 \tag{1.3.2}$$

where, for simplicity, $q_{N+1} = q_1$ (periodic boundary condition).

In this case there exist a large number of constants of motion, namely N:

1.3.3
$$A_k = (\underline{p} \cdot \underline{\eta}_k)^2 + \omega(k)^2(\underline{q} \cdot \underline{\eta}_k)^2 \quad k = 1, 2, \ldots, N \tag{1.3.3}$$

where $\underline{\eta}_1$, $\underline{\eta}_2, \ldots, \underline{\eta}_N$ are N suitable orthonormal vectors (*normal modes*) and $\omega(k)$ are the "*intrinsic pulsations*" of the chain:

1.3.4
$$\omega(k)^2 = 2(1 - \cos 2\pi k/N)\,. \tag{1.3.4}$$

The constants of motion in (1.3.3) can be arranged into an N-vector $\underline{A}(\Delta) = (A_1(\Delta), A_2(\Delta), \ldots, A_N(\Delta))$. The phase space cells Δ and Δ' for which the vectors $\underline{A}(\Delta)$ and $\underline{A}(\Delta')$ do not coincide cannot belong to the same cycle so that the system is not ergodic.

Nevertheless Boltzmann thought that circumstances like this should be considered exceptional. Hence it will be convenient not to go immediately into a deeper analysis of the ergodic problem: not only because of its difficulty but *mainly* because it is more urgent to see how one can proceed in the foundations of classical statistical mechanics.

Given a mechanical system of N identical (for the sake of simplicity) particles consider the problem of studying a fixed *observable* $f(\underline{p}, \underline{q})$ defined on phase space.

The first important quantity that one can study, and often the only one that it is necessary to study, is the *average value* of f:

1.3.5
$$\overline{f(\Delta)} = \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} f(S^k\Delta) \tag{1.3.5}$$

where $f(\Delta) = f(\underline{p},\underline{q})$ if $(\underline{p},\underline{q})$ is a point determining the cell Δ. If $\Delta_1 = \Delta, \Delta_2, \ldots, \Delta_{\mathcal{N}}$ is the cycle to which the cell Δ belongs, then:

1.3.6 $$\overline{f(\Delta)} = \frac{1}{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} f(\Delta_k) \qquad (1.3.6)$$

and in the ergodic case the cycle consists of the set of all cells with the same energy as Δ.

If the system energy is determined up to a macroscopic error DE, macroscopically negligible (but large with respect to the microscopic indetermination of energy δE, (1.2.14)), the cells with energy between $E - DE$ and E will be divided into cycles with (slightly) different energies. On each of the cycles the function f can be supposed to have the "*continuity*" property of having the same average value (*i.e.* energy independent up to negligible variations).

Hence, denoting by the symbol J_E the domain of the variables $(\underline{p},\underline{q})$ where $E - DE \le E(\underline{p},\underline{q}) \le E$ holds, one finds:

1.3.7 $$\overline{f(\Delta)} = \int_{J_E} f(\underline{p},\underline{q})\; \mathrm{d}p\; \mathrm{d}q \Big/ \int_{J_E} d\underline{p}\, d\underline{q}\,. \qquad (1.3.7)$$

Recalling in fact that the cells all have the same volume, (1.3.7) follows immediately from (1.3.6) and from the assumed negligibility of the dependence of $\overline{f(\Delta)}$ from $E(\Delta)$, *provided* h is so small that the sum over the cells can be replaced by an integral.

The above relation, which Boltzmann conjectured ([Bo71b], [Bo84]) to be always valid "discarding exceptional cases" (like the harmonic oscillator chain just described) and wrote in the suggestive form, [Bo71b], and p. 25 in [EE11]:

1.3.8 $$\lim_{T\to\infty} \frac{dt}{T} = \frac{d\underline{p}d\underline{q}}{\int_{J_E} d\underline{p}d\underline{q}} \qquad (1.3.8)$$

is read "the time average of an observable equals its average on the surface of constant energy". As we shall see, (§1.6), (1.3.8) provides a heuristic basis of the *microcanonical model* for classical thermodynamics.

Note that if (1.3.8) holds, *i.e.* if (1.3.7) holds, the average value of an observable will depend only upon E and not on the particular phase space cell Δ in which the system is found initially. The latter property is certainly a prerequisite that any theory aiming at deducing macroscopic properties of matter from the atomic hypothesis must possess.

It is, in fact, obvious that such properties cannot depend on the detailed microscopic properties of the configuration Δ in which the system happens to be at the initial time of our observations.

It is also relevant to note that in (1.3.6) the microscopic dynamics has disappeared: it is in fact implicit in the phase space cell enumeration, made so that $\Delta_1, \Delta_2, \Delta_3, \ldots$ are the cells into which Δ successively evolves at time

intervals τ. But it is clear that in (1.3.6) the order of such enumeration is not important and the same result would follow if the phase space cells with the same energy were enumerated differently.

Hence we can appreciate the fascination that the ergodic hypothesis exercises in *apparently* freeing us from the necessity of knowing the details of the microscopic dynamics, at least for the purposes of computing the observables averages. That this turns out to be an illusion, already clear to Boltzmann, see for instance p. 206 in [Bo74], will emerge from the analysis carried out in the following sections.

§1.4. Recurrence Times and Macroscopic Observables

In applications it has always been of great importance to be able to estimate the rapidity at which the limit $\overline{f}$ is reached: in order that (1.3.7) be useful it is necessary that the limit in (1.3.5) be attained within a time interval t which might be long compared to the microscopic τ but which should still be *very short* compared to the time intervals relevant for macroscopic observations that one wants to make on the system. It is, in fact, only on scales of the order of the macroscopic times t that the observable f may appear as constant and equal to its average value.

It is perfectly possible to conceive of a situation in which the system is ergodic, but the value $f(S^k\Delta)$ is ever changing, along the trajectories, so that the average value of f is reached on time scales of the order of magnitude of the time necessary to visit *the entire* surface of constant energy. The latter is necessarily enormous.

For instance, referring to the orders of magnitude discussed at the end of §1.2, see the values of $\delta p, \delta E$ preceding (1.2.16) and (1.2.16) itself, we can estimate this time by computing the number of cells with volume h^{3N} contained in the region between E and $E+\delta E$ and then multiplying the result by the characteristic time h/k_BT in (1.2.16), [Bo96], [Bo97].

If the surface of the d-dimensional unit sphere is written $2\sqrt{\pi}^d\Gamma(d/2)^{-1}$ (with Γ Euler's Gamma function) then the volume of the mentioned region, if h is very small, can be computed by using polar coordinates in momentum space. The cells are those such that $P \equiv \sqrt{\Sigma_i \underline{P}_i^2}$ varies between $P=\sqrt{2mE}$ and $P+\delta P = \sqrt{2m(E+\delta E)}$; hence we introduce, see §1.2, (1.2.14), the quantities:

1.4.1
$$\begin{aligned}
P =&\sqrt{2mE}, && \delta p = \overline{p} = \sqrt{3mk_BT}\\
\delta E =&3k_BT = \overline{p}\delta p/m, && \frac{E}{N} = \frac{3k_BT}{2}\\
\delta P =& \frac{\overline{p}\delta p}{P} = \frac{\overline{p}}{\sqrt{N}}, && \delta q = \Big(\frac{V}{N}\Big)^{1/3}
\end{aligned} \tag{1.4.1}$$

where k_B is Boltzmann's constant, $k_B = 1.38\times 10^{-16}\, erg^\circ K^{-1}$, T is the absolute temperature, V is the volume occupied by the gas and N is the

particle number. One finds that the volume we are trying to estimate is, setting $h = \delta p \delta q$ and using Stirling's formula to evaluate $\Gamma(\frac{N}{2})$:

1.4.2
$$\begin{aligned} w = & V^N \sqrt{2mE}^{3N-1} \, \delta P \, \sqrt{\pi}^{3N} 2\Gamma(3N/2)^{-1} = \\ & = (N\delta q^3)^N (\sqrt{N}\delta p)^{3N-1} \delta P \sqrt{\pi}^{3N} 2\Gamma(3N/2)^{-1} = \\ & = (\delta p \delta q)^{3N} N^{\frac{5}{2}N-1} \sqrt{\pi}^{3N} 2\Gamma(3N/2)^{-1} \cong \\ & \cong h^{3N} \frac{N^{\frac{5}{2}N-1}}{N^{\frac{3}{2}N-\frac{1}{2}}} \big(\frac{2\pi e}{3}\big)^{3N/2} \frac{2}{\sqrt{\pi}} \cong h^{3N} N^{N-\frac{1}{2}} \Big(\frac{2\pi e}{3}\Big)^{3N/2} . \end{aligned} \tag{1.4.2}$$

The number of cells $\mathcal{N}$ would then be w/h^{3N}. But we shall assume, on the grounds of particle indistinguishability, that cells differing because individual particles are permuted are in fact identical. Then (1.4.2) has to be divided by $N! \cong N^N e^{-N}\sqrt{2\pi N}$ and therefore the *recurrence time*, if the system did move ergodically on the surface of energy E, would be:

1.4.3
$$T_{\rm recurrence} = \mathcal{N}\tau \cong \mathcal{N}\frac{h}{k_B T} \cong N^{-1}\frac{h}{k_B T}\left(\frac{2\pi e^{5/3}}{3}\right)^{3N/2} . \tag{1.4.3}$$

As discussed in §1.2 the order of magnitude of $\tau = h/k_B T$ is, if $T = 300°K$, of about 10^{-14} *sec*. For our present purposes it makes no difference whether we use the expression $h = \delta p \delta q$ with $\delta p, \delta q$ given in (1.4.1) with $V = 1\ {\rm cm}^3$, $N = 2.7 \times 10^{19}$, $m = 3.34 \times 10^{-24}\, g$ = hydrogen molecule mass, or whether we use Planck's constant (see comment after (1.2.18)).

Hence, $\frac{2\pi}{3}e^{\frac{5}{/}3}$ being > 10, the recurrence time in (1.4.3) is unimaginably longer than the age of the Universe as soon as N reaches a few decades (still very small compared to Avogadro's number). If T is chosen to be 0℃: for $1\, cm^3$ of hydrogen at $0°C$, $1\, atm$ one has $N \simeq 10^{19}$ and $T_{\rm recurrence} = 10^{-14} \cdot 10^{10^{19}}\, sec$, while the age of the Universe is *only* $\sim 10^{17}\, sec$!

Boltzmann's idea to reconcile ergodicity with the observed rapidity of the approach to equilibrium was that the interesting observables, the *macroscopic observables*, had an essentially constant value on the surface of given energy *with the exception of an extremely small fraction* ε *of the cells*, [Bo74], p. 206. See §1.7 below for further comments.

Hence the time necessary to attain the asymptotic average value will not be of the order of magnitude of the hyperastronomic recurrence time, but rather of the order of $T' = \varepsilon T_{recurrence}$. And one should think that $\varepsilon \to 0$ as the number of particles grows and that T' is very many orders of magnitude smaller than T so that it becomes observable on "human" time scales, see §1.8 for a quantitative discussion (actually T' sets, essentially by definition, the size of the human time scale).

Examples of important macroscopic observables are:

•(1) the ratio between the number of particles located in a small cube Q and the volume of Q: this is an observable that will be denoted $\rho(Q)$ and its average value has the interpretation of *density* in Q;

•(2) the sum of the kinetic energies of the particles: $K(\Delta) = \Sigma_i \underline{p}_i^2/2m$;

•(3) the total potential energy of the system: $\Phi(q) = \Sigma_{i<j}\varphi(\underline{q}_i - \underline{q}_j)$;

•(4) the number of particles in a small cube Q adherent to the container walls, and having a negative component of velocity along the inner normal with value in $[-v, -(v+dv)]$, $v > 0$. This number divided by the volume of Q is the "density" $n(Q, v)dv$ of particles with normal velocity $-v$ that are about to collide with the external walls of Q. Such particles will cede a momentum $2mv$ normally to the wall at the moment of their collision (as their momentum will change from $-mv$ to mv) with the wall. Consider the observable defined by the sum over the values of v and over the cubes Q adjacent to the boundary of the container V:

1.4.4
$$\sum_Q \int_{v>0} d\,v\,n(Q,v)(2mv)\frac{vs}{S} \;=\; P(\Delta) \qquad (1.4.4)$$

with s = area of a face of Q and S = area of container surface: this is the momentum transferred, per unit time and surface area, to the wall (note that the number of collisions with the wall per unit time on the face s of Q adjacent to the walls and with normal velocity v is $n(Q, v)vs\,\mathrm{d}v$). The quantity (1.4.4) is an observable (*i.e.* a function on phase space) whose average value has the interpretation of macroscopic *pressure*, therefore it can be called the "microscopic pressure" in the phase space point $(\underline{p}, \underline{q})$.

•(5) the product $\rho(Q)\rho(Q')$ is also interesting and its average value is called the density *pair correlation function* between the cubes Q, Q'. Its average value provides information on the joint probability of finding simultaneously a particle in Q and one in Q'.

§1.5. Statistical Ensembles or "Monodes" and Models of Thermodynamics. Thermodynamics without Dynamics

From a more general viewpoint and without assuming the ergodic hypothesis it is clear that the average value of an observable will always exist and it will be equal to its average over the cycle containing the initial datum, see (1.3.6).

For a more quantitative formulation of this remark we introduce the notion of *stationary distribution*: it is a function associating with each phase space cell a number $\mu(\Delta)$ (*probability* or *measure* of Δ) so that:

1.5.1
$$\mu(\Delta) \ge 0 \qquad \sum_\Delta \mu(\Delta) = 1 \qquad \mu(\Delta) = \mu(S\Delta) \qquad (1.5.1)$$

if S is the time evolution map which permutes the cells, see §1.2 and §1.3.

One usually says that μ is an *invariant probability distribution* or a *stationary probability distribution* on phase space (or, better, on phase space cells). The following definition will be convenient (see §1.3):

Definition: Let μ be an invariant probability distribution on phase space cells. If the dynamics map S acts as a one-cycle permutation of the set of cells Δ for which $\mu(\Delta) > 0$ then μ is called ergodic.

If one imagines covering phase space with a fluid so that the fluid mass in Δ is $\mu(\Delta)$ and if the phase space point are moved by the permutation S associated with the dynamics then the fluid looks immobile, *i.e.* its distribution on phase space remains invariant (or stationary) as time goes by: this gives motivation for the name used for μ.

It is clear that $\mu(\Delta)$ must have the same value on all cells belonging to the same cycle $\mathcal{C}_\alpha$ of the permutation S (here α is a label distinguishing the various cycles of S). If $\mathcal{N}(\mathcal{C}_\alpha)$ is the number of cells in the cycle $\mathcal{C}_\alpha$ it must, therefore, be that $\mu(\Delta) = p_\alpha/\mathcal{N}(\mathcal{C}_\alpha)$, with $p_\alpha \ge 0$ and $\sum_\alpha p_\alpha = 1$, for $\Delta \in \mathcal{C}_\alpha$.

It is useful to define, for each cycle $\mathcal{C}_\alpha$ of S, a (ergodic) stationary distribution μ_α by setting:

1.5.2 $$\mu_\alpha(\Delta) = \begin{cases} 1/\mathcal{N}(\mathcal{C}_\alpha) & \text{if } \Delta \in \mathcal{C}_\alpha \\ 0 & \text{otherwise} \end{cases} \tag{1.5.2}$$

and this allows us to think that any invariant probability distribution is a linear combination of the μ_α associated with the various cycles of S:

1.5.3 $$\mu(\Delta) = \sum_\alpha p_\alpha \mu_\alpha(\Delta)\,, \tag{1.5.3}$$

where $p_\alpha \ge 0$ are suitable coefficients with $\sum_\alpha p_\alpha = 1$, which can be called the "probabilities of the cycles" in the distribution μ. Note that, by definition, each of the distributions μ_α is ergodic because it gives a positive probability only to cells that are part of the same cycle (namely $\mathcal{C}_\alpha$).

The decomposition (1.5.3) of the most general S-invariant distribution μ as a sum of S-ergodic distributions is naturally called the *ergodic decomposition* of μ (with respect to the dynamics S).

In the deep paper [Bo84] Boltzmann formulated the hypothesis that stationary distributions μ could be interpreted as macroscopic equilibrium states so that the set of macroscopic equilibrium states could be identified with a subset $\mathcal{E}$ of the stationary distributions on phase space cells. The current terminology refers to this concept as an *ensemble*, after Gibbs: while Boltzmann used the word *monode*. We shall call it an *ensemble* or a *statistical ensemble*.

Identification between an individual stationary probability distribution μ on phase space and a corresponding macroscopic equilibrium state takes place by identifying $\mu(\Delta)$ with the probability of finding the system in the cell (*i.e.* in the microscopic state) Δ if one performed, at a randomly chosen time, the observation of the microscopic state.

Therefore the average value in time, in the macroscopic equilibrium state described by μ, of a generic observable f would be:

1.5.4 $$\overline{f} = \sum_\Delta \mu(\Delta) f(\Delta)\,. \tag{1.5.4}$$

This relation correctly gives, in principle, the average value of f in time, if the initial data are chosen randomly with a distribution μ which is ergodic. But in general *even if μ is ergodic* one should not think that (1.5.4) is directly related to the physical properties of μ. This was already becoming clear in §1.3 and §1.4 when we referred to the length of the recurrence times and hence to the necessity of further assumptions to derive (1.3.7), (1.3.8).

We shall come back to (1.5.4) and to the ergodic hypothesis in §1.6. Returning to Boltzmann's statistical ensembles he raised the following questionin in the paper [Bo84]: letting aside the ergodic hypothesis or any other attempt at a dynamical justification of (1.5.4), consider all possible statistical ensembles $\mathcal{E}$ of stationary distributions on phase space. Fix $\mathcal{E}$ and, for each $\mu \in \mathcal{E}$, define:

1.5.5

$$\begin{aligned}
\Phi(\mu) = \sum_\Delta \mu(\Delta)\Phi(\Delta) &= \qquad \text{“average potential energy”},\\
T(\mu) = \sum_\Delta \mu(\Delta)K(\Delta) &= \qquad \text{“average kinetic energy”},\\
U(\mu) = T(\mu) + \Phi(\mu) &= \qquad \text{“average total energy”}, \qquad (1.5.5)\\
P(\mu) = \sum_\Delta \mu(\Delta)P(\Delta) &= \qquad \text{“pressure”, see (1.4.4)},\\
\rho(\mu) = N/V = \rho \equiv 1/v &= \qquad \text{“density”},\\
V = \int \mathrm{d}q &= \qquad \text{“volume”},
\end{aligned}$$

where V is the volume assigned to the system (*i.e.* the volume of the container) and N is the particle number.

Question (“orthodicity problem”): which statistical ensembles, or monodes, $\mathcal{E}$ have the property that as μ changes infinitesimally within $\mathcal{E}$ the corresponding infinitesimal variations dU, dV of $U = U(\mu)$ and V, see (1.5.5), are related to the pressure $p = P(\mu)$ and to the average kinetic energy per particle $\overline{T} = T(\mu)/N$) so that:

1.5.6

$$\frac{dU + P\,dV}{\overline{T}} = \textit{exact differential} \qquad (1.5.6)$$

at least in the thermodynamic limit in which the volume $V \to \infty$ and also $N, U \to \infty$ so that the densities N/V, U/V remain constant (assuming for simplicity that the container keeps cubic shape).

Ensembles (or monodes) satisfying the property (1.5.6) were called by Boltzmann *orthodes*: they are, in other words, the statistical ensembles $\mathcal{E}$ in which it is possible to interpret the average kinetic energy per particle, $\overline{T}$, as proportional to the absolute temperature T (via a proportionality constant, to be determined empirically and conventionally denoted $(2/3k_B)^{-1}$: so that $T = \frac{2}{3k_B}\frac{T(\mu)}{N}$); and furthermore it is possible to define via (1.5.6) a

function $S(\mu)$ on $\mathcal{E}$ so that the observables U, ρ, T, V, P, S satisfy the relations that the classical thermodynamics quantities with the corresponding name satisfy, at least in the thermodynamic limit.

In this identification the function $S(\mu)$ would become, naturally, the *entropy* and the validity of (1.5.6) would be called the *second law*.

In other words Boltzmann posed the question of when it would be possible to interpret the elements μ of a statistical ensemble $\mathcal{E}$ of stationary distributions on phase space as macroscopic states of a system governed by the laws of classical thermodynamics.

The ergodic hypothesis combined with the other assumptions used in §1.3 to deduce (1.3.7), (1.3.8) leads us to think that the statistical ensemble $\mathcal{E}$ consisting of the distributions μ on phase space defined by:

1.5.7
$$\begin{aligned} &\mu(\Delta) = 1/\mathcal{N}(U,V) \qquad && \text{if} \quad E(\Delta) \in (U - DE, U) \\ &\mu(\Delta) = 0 && \text{otherwise} \end{aligned} \tag{1.5.7}$$

where U, V are prefixed parameters corresponding to the total energy and volume of the system, should *necessarily* be a statistical ensemble apt at describing the macroscopic equilibrium states. Here $\mathcal{N}(U,V)$ is a normalization constant to be identified as proportional to the integral $\int d\underline{p}d\underline{q}$ over the region J_E of $\underline{p}$, $\underline{q}$ in which $E(\underline{p},\underline{q}) \in (U - DE, U)$; and the parameter DE is "arbitrary" as discussed before (1.3.7).

However the orthodicity or nonorthodicity of a statistical ensemble $\mathcal{E}$ whose elements are parameterized by U, V as in (1.5.7) is "only" the question of whether (1.5.6) (second law) holds or not and this problem is not, in itself, logically or mathematically related to any microscopic dynamics property.

The relation between orthodicity of a statistical ensemble and the hypotheses on microscopic dynamics (like the ergodic hypothesis) that would *a priori* guarantee the physical validity of the ensuing model of thermodynamics will be reexamined in more detail at the end of §1.6.

If there were several orthodic statistical ensembles then each of them would provide us with a mechanical microscopic *model of thermodynamics*: of course if there were several possible models of thermodynamics (*i.e.* several orthodic statistical ensembles) it *should* also happen that they give *equivalent* descriptions, *i.e.* that they give the *same expression* to the entropy S as a function of the other thermodynamic quantities, so that thermodynamics would be described in mechanical terms in a nonambiguous way. This check is therefore one of the main tasks of the statistical ensembles theory.

It appears that in attempting to abandon the (hard) fundamental aim at founding thermodynamics on microscopic dynamics one shall nevertheless not avoid having to attack difficult questions like that of the nonambiguity of the thermodynamics that corresponds to a given system. The latter is a problem that has been studied and solved in various important cases, but we are far from being sure that such cases (the *microcanonical* or the *canonical* or the *grand canonical* ensembles to be discussed below, and others) exhaust all possible ones. Hence a "complete" understanding of this question could

reveal itself equivalent to the dynamical foundation of thermodynamics: the very problem that one is hoping to circumvent by deciding to "only" build a mechanical model of thermodynamics, *i.e.* an orthodic ensemble.

§1.6. Models of Thermodynamics. Microcanonical & Canonical Ensembles and the Ergodic Hypothesis.

The problem of the existence of statistical ensembles (*i.e.* a family of stationary probability distributions on phase space) that provides mechanical models of thermodynamics[2] was solved by Boltzmann in the same paper quoted above, [Bo84] (following earlier basic papers on the canonical ensemble [Bo71a], [Bo71b] where the notion of ensemble seems to appear for the first time).

Here Boltzmann showed that the statistical ensembles described below and called, after Gibbs, the *microcanonical* and the *canonical ensemble* are *orthodic*, *i.e.* they define a microscopic model of thermodynamics in which the average kinetic energy per particle is proportional to absolute temperature (see below and §1.5).

(1) *The microcanonical ensemble*

It was named in this way by Gibbs while Boltzmann referred to it by the still famous, but never used, name of *ergode*. The microcanonical ensemble consists in the collection $\mathcal{E}$ of stationary distributions μ parameterized by two parameters $U=$ total energy and $V=$ system volume so that, see (1.5.2):

1.6.1
$$\begin{aligned} &\mu(\Delta) = 1/\mathcal{N}(U,V) \qquad && \text{if } U - DE \le E(\Delta) \le U \\ &\mu(\Delta) = 0 && \text{otherwise} \end{aligned} \tag{1.6.1}$$

with:

1.6.2
$$\mathcal{N}(U,V) = \sum_{U-DE\le E(\Delta)\le U} 1 = \{\text{number of cells } \Delta \text{ with energy } E(\Delta) \in (U-DE, U)\} \tag{1.6.2}$$

where the quantity DE has to be a quantity, possibly V-dependent, "macroscopically negligible" compared to U, such that one may think that all cells with energy between $U-DE$ and U have the "same energy" from a macroscopic point of view.

The importance of the microcanonical ensemble in the relation between classical thermodynamics and the atomic hypothesis is illustrated by the argument leading to (1.3.8) which proposes it as the natural candidate for an example of an orthodic ensemble.

---

[2] At least in the thermodynamic limit, see (1.5.6), in which the volume becomes infinite but the average density and energy per particle stay fixed.

However, as discussed in §1.5, the argument leading to (1.3.7), (1.3.8) cannot possibly be regarded as a "proof on physical grounds" of orthodicity of the microcanonical ensemble.[3]

Following the general definition in §1.5 of orthodic statistical ensemble, *i.e.* of an ensemble generating a model of thermodynamics, we can define the "absolute temperature" and the "entropy" of every element μ ("macroscopic state") so that the temperature T is proportional to the average kinetic energy. Boltzmann showed that such functions T and S are given by the celebrated relations:

1.6.3 $$T = \frac{2}{3k_B}\frac{T(\mu)}{N}\,, \qquad S(\mu) = k_B \log \mathcal{N}(U,V) \tag{1.6.3}$$

where k_B, "Boltzmann's constant", is a universal constant to be empirically determined by comparison between theory and experiment.[4] The factor $\frac{2}{3}$ is conventional and its choice simplifies some of the following formulae, besides the second of (1.6.3).[5]

The statement that (1.6.1), (1.6.2) provide us with a microscopic *model of thermodynamics* in the thermodynamic limit $V\to\infty$, $U\to\infty$, $N\to\infty$ so that $u=U/N$, $v=V/N$ remain constant has to be interpreted as follows.

One evaluates, starting from (1.6.1)÷(1.6.3) (see also (1.5.5)):

1.6.4 $$\begin{aligned} &u = U/N = \text{ “specific energy” }, \ v = V/N = \text{ “specific volume”},\\ &T = 2T(\mu)/3k_BN = \text{ “temperature” },\\ &s = S(\mu)/N = \text{ “entropy”}, \, , \ p = P(\mu) = \text{ “pressure”}. \end{aligned} \tag{1.6.4}$$

Since the quantities u, v determine $\mu\in\mathcal{E}$ it will be possible to express T,p,s in terms of u, v via functions $T(u,v)$, $P(u,v)$, $s(u,v)$ that we shall suppose to admit a limit value in the thermodynamic limit (*i.e.* $V\to\infty$ with fixed u, v).

Then to say that (1.6.1), (1.6.2) give a *model of thermodynamics* means (see also §1.5) that such functions satisfy the same relations that link the quantities with the same name in classical thermodynamics, namely:

1.6.5 $$\mathrm{d}u = T\,\mathrm{d}s - P\,\mathrm{d}v\,. \tag{1.6.5}$$

Equation (1.6.5) is read as follows: if the state μ defined by (1.6.1),(1.6.2) is subject to a small variation by changing the parameters U,V that define it, then the corresponding variations of u, s, v verify (1.6.5), *i.e.* the second principle of thermodynamics: see Chap.II for a discussion and a proof of (1.6.3), (1.6.5).

The proof of a statement like (1.6.5) for the ensemble $\mathcal{E}$ was called, by Boltzmann, a proof of the *heat theorem*.

---

[3] Which it is worth stressing once more does not depend on the microscopic dynamics.

[4] As already said and as it will be discussed later, one finds $k_B = 1.38^{-16}\, erg°K^{-1}$.

[5] Mainly it simplifies the relation between T and β in the first of (1.6.8) below.

(2) *The canonical ensemble.*

The name was introduced by Gibbs, while Boltzmann referred to it with the name of *holode*. It consists in the collection $\mathcal{E}$ of stationary distributions μ parameterized by two parameters β and $v = V/N$, via the definition:

1.6.6
$$\mu(\Delta) = (\exp -\beta E(\Delta))/Z(\beta, V) \tag{1.6.6}$$

with

1.6.7
$$Z(\beta, V) = \sum_{\Delta} \exp -\beta E(\Delta)\,. \tag{1.6.7}$$

Boltzmann proved the proportionality between $T(\mu)$ and β^{-1} as well as the orthodicity of this statistical ensemble by showing that temperature and entropy can be defined by

1.6.8
$$T = 2T(\mu)/3k_B N = 1/k_B\beta \qquad S = -k_B(\beta U - \log Z(\beta, V)) \tag{1.6.8}$$

where k_B is a universal constant to be empirically determined.

The statement that (1.6.6), (1.6.8) provide us with a model of thermodynamics, in the thermodynamic limit $V \to \infty$, $V/N \to v$, $\beta =$ constant, has the same meaning discussed in the previous case of the microcanonical ensemble. See Chap.II for the analysis of the orthodicity of the canonical ensemble, *i.e.* for a proof of the heat theorem for the canonical ensemble.

The relations (1.6.5) hold, as already pointed out, for both ensembles considered, hence each of them gives a microscopic "mechanical" model of classical thermodynamics.

Since entropy, pressure, temperature, *etc*, are in both cases explicitly expressible in terms of two independent parameters (u, v or β, v) it will be possible to compute the *equation of state* (*i.e.* the relation between p, v and T) in terms of the microscopic properties of the system, at least in principle: this is enormous progress with respect to classical thermodynamics where the equation of state always has a phenomenological character, *i.e.* it is a relation that can only be deduced by means of experiments.

It is clear, however, that the models of thermodynamics described above must respond, to be acceptable as physical theories, to the basic prerequisite of defining not only a possible thermodynamics[6] but also of defining *the* thermodynamics of the system, which is experimentally accessible. One can call the check of the two prerequisites a check of theoretical and experimental consistency, respectively.

For this it is necessary, first, that the two models of thermodynamics coincide (*i.e.* lead to the same relations between the basic thermodynamic quantities u, v, T, P, s) but it is also necessary that the two models agree with the experimental observations.

---

[6] *I.e* a thermodynamics that does not come into conflict with the basic principles, expressed by (1.6.5).

But *a priori* there are no reasons that imply that the above two prerequisites hold.

Here it is worthwhile to get more deeply into the questions we raised in connection with (1.3.7) and to attempt a justification of the validity of the microcanonical ensemble as a model for thermodynamics with physically acceptable consequences and predictions. This leads us once more to discuss the ergodic hypothesis that is sometimes invoked at this point to guarantee *a priori* or to explain *a posteriori* the success of theoretical and experimental consistency checks, whose necessity has been just pointed out.

In §1.3 we have seen how the microcanonical distribution could be justified as describing macroscopic equilibrium states on the basis of the ergodic hypothesis and of a continuity property of the averages of the relevant observables (see the lines preceding (1.3.7)): in that analysis, leading to (1.3.7), we have not taken into account the time scales involved. Their utmost importance has been stated in §1.4: if (1.3.7) held but the average value over time of the observable f, given by the right-hand side of (1.3.7), was attained in a hyperastronomic time, comparable to the one given by (1.4.3), then (1.3.7) would, obviously, have little practical interest and value.

§1.7. Critique of the Ergodic Hypothesis

Summarizing: to deduce (1.3.7), hence for an *a priori* justification of the connection between the microcanonical ensemble and the set of states of macroscopic thermodynamic equilibrium, one meets three main difficulties.

• The first is a verification of the ergodic hypothesis, §1.3, as a mathematical problem.

• The second is that even accepting the ergodic hypothesis for the cyclicity of the dynamics on the surface with constant energy (*i.e.* with energy fixed within microscopic uncertainty δE) one has to solve the difficulty that, in spite of the ergodicity, the elements of the microcanonical ensemble are not ergodic *because* the (trivial) non ergodicity is due to the fact that in the microcanonical ensemble the energy varies by a *small but macroscopic quantity* $DE \gg \delta E$.

• The third is that, in any event, it would seem that enormous times are needed before the fluctuations of the time averages over finite times stabilize around the equilibrium limit value (times enormously longer than the age of the Universe).

The three difficulties would be solved if one supposed that, simultaneously:

(i) the phase space cells with fixed energy (microscopically fixed) are part of a single cycle of the dynamics S: this is the ergodic hypothesis, see §1.3.

(ii) the values of the "relevant" macroscopic observables are essentially the same on cells corresponding to a given macroscopic value of the energy

E, with the possible exception of a small fraction of the number of cells, negligible for large systems *i.e.* in the thermodynamic limit.

(iii) the common average value that the relevant observables assume on trajectories of cells of energy E changes only slightly as the total energy changes between the values U and $U-DE$, if U and DE are two macroscopic values with $U \gg DE$ (but $DE \gg \delta E$). This can be called a *continuity assumption.*

The hypotheses (i) and (iii), see §1.3, show that the average values of the macroscopic observables can be computed by using, equivalently, any ergodic component of a given microcanonical distribution μ.

Hypothesis (ii) allows us to say that the time necessary in order that an average value of an observable be attained, if computed on the evolution of a particular microscopic state Δ, is by far shorter than the recurrence time (too long to be interesting or relevant). The region of phase space where macroscopic observables take the equilibrium value sometimes has been pictorially called the "Boltzmann's sea" (see [Bo74], p. 206, and [Ul68], p. 3, fig. 2).

Accepting (i), (ii) and (iii) implies (by the physical meaning that u, p, v acquire) that the microcanonical ensemble *must* provide a model for thermodynamics in the sense that $dU + p\,dV$ must admit an integrating factor (to be identified with the absolute temperature). The fact that this factor turns out to be proportional to the average kinetic energy is, from this viewpoint (*and only in the case of classical statistical mechanics* as one should always keep in mind), a consequence (as we shall show in Chap.II).

One can remark that assumptions (ii) and (iii) are assumptions that do not involve explicitly the dynamical properties, at least on a qualitative level: one says that they are *equilibrium properties* of the system. And it is quite reasonable to think that they are satisfied for the vast majority of systems encountered in applications, because in many cases it is possible to really verify them, sometimes even with complete mathematical rigor, [Fi64], [Ru69].

Hence the deeper assumption is in (i), and it is for this reason that sometimes, quite improperly, it is claimed that the ergodic hypothesis is "*the theoretical foundation for using the microcanonical ensemble as a model for the equilibrium states of a system*".

The improper nature of the above locution lies in the fact that (i) can be *greatly weakened* without leading to a modification of the inferences on the microcanonical ensemble.

For instance one could simply require that *only* the time average of few macroscopically interesting observables should have the same value on every cycle (or on the great majority of cycles) of the dynamics with a fixed energy.

This can be done while accepting the possibility of many different cycles (on which non macroscopically interesting observables would take different average values). An essentially exhaustive list of the "few" interesting

observables for monoatomic gases is given by (1.5.5).

Furthermore the above-mentioned locution is improper also because, even if one accepts it, it cannot release us from checking (ii), (iii) which, in particular, require a quantitative verification: evidently one cannot be satisfied with a simple qualitative verification since the orders of magnitude involved are very different. One could, in fact, raise doubts that the time "for reaching equilibrium" could really come down from the recurrence times (superastronomical) to the times experimentally recorded (usually of a few microseconds).

For what concerns the canonical ensemble its use could be justified simply by proving that it leads to the same results that one obtains by using the microcanonical ensemble, at least in the thermodynamic limit and for the few interesting observables (see above for a list).

But, as already mentioned, the ergodic hypothesis (with or without the extra two assumptions (ii), (iii) above) is technically too difficult to study and for this reason an attempt has been made to construct models of thermodynamics while avoiding solving, even if partially, the ergodic problem.

The proposal is simply to prove that all the orthodic ensembles (at least the reasonable ones)[7] generate the *same* macroscopic thermodynamics (for instance the same equation of state). This property, by itself very notable and remarkable, should then be considered sufficient to postulate, by the "principle of sufficient reason", that the equations of state of a system can be calculated from the microscopic properties (*i.e.* from the Hamiltonian of the system) by evaluating the average values of the basic observables (see (1.5.5)) via the distributions of the microcanonical or canonical ensembles, or more generally of any orthodic ensemble.

The latter is the point of view usually attributed to Gibbs: virtually all the treatises on statistical mechanics are based on it.

It is well understandable why such a point of view appeared unsatisfactory to Boltzmann who had the ambition of reducing thermodynamics to mechanics without introducing any new postulate: on the other hand, the pragmatic approach of Gibbs is also very understandable if one keeps in

---

[7] One should not think that it is difficult to devise ensembles which are orthodic and which may seem "not reasonable" (for a thermodynamic interpretation): in fact Boltzmann's paper, [Bo84], on the ensembles *starts* with such an example involving the motion of one of Saturn's rings regarded as a massive line (in a parallel paper the example was the Moon, whose orbit was replaced by an ellipse of mass such that each arc contained an amount of mass proportional to the time spent on it by the Moon). This may have been one of the reasons this fundamental paper has been overlooked for so many years. Such "unphysical" examples come from Helmoltz, [He95a], [He95b], and played an important role for Boltzmann (who was considering them in a less systematic way even much earlier, [Bo66]). In fact if one can define the mechanical analogue of thermodynamics for *any* system, small or large, then it is natural to think that in large systems the average quantities will also satisfy the second law. And the idea (of Boltzmann) that the macroscopic observables have the same value on most of the energyy surface makes the law easily observable in large systems, while this may not be the case in very small systems. In other words the one-degree-of-freedom examples are not at all unphysical; rather the contrary holds: see Appendix 1.A1 (to Chap.I) for Helmoltz's theory and Chap.IX for a recent application of the same viewpoint.

mind the necessity of deducing all the applicative consequences stemming from the marvelous discovery of the possibility of *unambiguously deriving* values of thermodynamical quantities in terms of mechanical properties of the atomic model of matter.

For the past few decades, about a century after the birth of the above theories, we seem to feel again the necessity of a unified derivation of thermodynamics from mechanics without the artificial *a priori* postulate that thermodynamics is described by the orthodic statistical ensembles; a postulate made possible, *i.e.* consistent, by the mentioned independence (*discussed in the following Chap.II*) of the results as functions of the statistical ensemble used.

The ergodic problem and the statistical dynamics are therefore again at the center of research, and are stimulating new interesting ideas and results.

Boltzmann tried to justify the microcanonical and canonical ensembles also following a path rather different from the one of studying the ergodic problem and the hypotheses (i),(ii),(iii) above, [UF63]. And his attempt led him, [Bo72], to deduce the *Boltzmann's equation* which revealed itself essential even for technical applications, although it presented and presents various conceptually unsatisfactory aspects, see §1.8 for a first analysis of this equation.

§1.8. Approach to Equilibrium and Boltzmann's Equation. Ergodicity and Irreversibility

As discussed in the previous sections macroscopic equilibrium states can be identified with elements of the orthodic statistical ensembles (microcanonical, canonical, grand canonical, ...). It is not quantitatively clear, however, through which mechanism a mechanical system initially in non equilibrium can reach equilibrium.

We have argued that the ergodic hypothesis, by itself, is not sufficient to explain why a system reaches equilibrium within times usually relatively short.

Boltzmann developed a model, [Bo72], for describing approach to equilibrium which was strongly criticized since its formulation, much as his other intuitions, and which is considered by some (perhaps incorrectly) his greatest contribution to Science.

The validity of the model is limited to systems with so low a density that they can be considered rarefied gases and this shows how it can, in concrete cases, happen that assumptions (i), (ii), (iii) of §1.7 could be, for practical purposes, verified in such systems and how it could be possible that the interesting observables reach their average values over time scales accessible to our senses rather than on the absurdly long recurrence time scales.

One imagines the system to consist of N identical particles (for simplicity), each of which is described by momentum $\underline{p}$ and by position $\underline{q}$. They move as if free, *except* that from time to time they collide.

Assuming that such particles are rigid spheres with radius R (again only

for simplicity) and that they have an average speed $\overline{v}$, the low-density assumption is that the density $\rho = N/V$ is such that

1.8.1
$$\rho R^3 \ll 1 \tag{1.8.1}$$

which means that it is very unlikely that there are two particles at a distance of the order R, *i.e.* "colliding".

At the same time one requires that the number of collisions that each particle undergoes per unit time does not vanish. Evidently this number has order of magnitude:

1.8.2
$$\rho R^2\,\overline{v}\,. \tag{1.8.2}$$

Hence the limit situation in which the gas is very rarefied but, nevertheless, the number of collisions that each particle undergoes per unit time is not negligible, is described by

1.8.3
$$\begin{aligned} R\to 0,\ \rho\to\infty \qquad\qquad & \text{so that}\\ \rho R^3\to 0,\ \rho R^2\,\overline{v} = w = \ & \text{fixed quantity}\,. \end{aligned} \tag{1.8.3}$$

The quantity $\tau = 1/w$ is the *time of flight* between two collisions while the *mean free path* is $\tau\,\overline{v} = 1/\rho R^2$.

The limit situation that is obtained by letting $R\to 0$ and $\rho\to\infty$ as in (1.8.3) is called *Grad's limit*. In the situation envisaged by Boltzmann one supposes that we are "close" to this limit, *i.e.* one supposes that we are close to and $\rho R^2\,\overline{v} = w > 0$.

It is of some interest to compute ρR^3, τ and $\tau\,\overline{v}$ for a Hydrogen sample at atmospheric pressure and room temperature ($p = 1\,atm, T = 293^o K$): one finds $\rho R^3 = 5.8\times 10^{-4}$, $\tau = 2.5\times 10^{-10}\,sec$, $\overline{v} = 1.9\times 10^3\,m/sec$.

Let then $f(\underline{p},\underline{q})\,\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}$ be the number of particles that can be found in the cell $Q = \mathrm{d}\underline{p}\,\mathrm{d}\underline{q}$ of the phase space describing the states of a *single* particle (not to be confused with the phase space which we have been using so far, which describes the states of N particles).

Boltzmann remarks that f can change in time either by virtue of collisions or because particles move in space. If ε is a prefixed time interval, the number of particles that at a certain instant are in the cell Q is:

1.8.4
$$\begin{aligned} & f(\underline{p},\underline{q},\underline{t})\,\mathrm{d}\underline{p}\,\mathrm{d}\underline{q} = f(\underline{p},\underline{q}-\varepsilon\underline{p}/m, t-\varepsilon)\,\mathrm{d}\underline{p}\,\mathrm{d}\underline{q} + \\ & + \sum_{Q',Q''} \begin{array}{l}\text{(number of particles in } Q' \text{ that collide per unit of time with}\\ \text{particles in } Q'' \text{ producing particles in } Q_1, Q_2 \text{ with } Q_1\equiv Q)\end{array} \\ & - \sum_{Q',Q''} \begin{array}{l}\text{(number of particles in } Q_1\equiv Q \text{ that collide per unit of time}\\ \text{with particles in } Q_2 \text{ producing particles in } Q', Q'')\end{array} \end{aligned} \tag{1.8.4}$$

If we consider the collision that transforms two particles in Q', Q'' into two others in Q_1, Q_2 we must have, by the conservation of momentum and energy in the collision):

1.8.5
$$\underline{p}' + \underline{p}'' = \underline{p}_1 + \underline{p}_2\,, \qquad \underline{p}'^2 + \underline{p}''^2 = \underline{p}_1^2 + \underline{p}_2^2 \tag{1.8.5}$$

and the number of collisions leading from $\underline{p}'$, $\underline{p}''$ to $\underline{p}_1$, $\underline{p}_2$ can be expressed in terms of the notion of collision *cross-section* $\sigma = \sigma(\underline{p}', \underline{p}''; \underline{p}, \underline{p}_2)$.

The latter is defined to be the fraction of particles of a stream with momentum in $d\underline{p}$ and spatial density $n(\underline{p})d\underline{p}$ streaming around *one particle* with momentum $\underline{p}_2$ that collides with it in time dt, experiencing a collision that trasnforms $\underline{p}, \underline{p}_2$ into $\underline{p}', \underline{p}''$. This number is written as $n(\underline{p})d\underline{p}\,\frac{|\underline{p}-\underline{p}_2|}{m}\sigma(\underline{p}', \underline{p}''; \underline{p}, \underline{p}_2)\,dt$ and σ has the dimension of a surface. Hence the total number of such collision per unit volume will be $n(\underline{p}_2)d\underline{p}_2\, n(\underline{p})d\underline{p}\,[\frac{|\underline{p}-\underline{p}_2|}{m}\sigma(\underline{p}', \underline{p}''; \underline{p}, \underline{p}_2)]\,dt$, *i.e.* the the number of particles with momentum in $d\underline{p}$ that experience a collision with one $\underline{p}_2$-particle in time dt is the number of particles with momentum in $d\underline{p}$ and contained in a volume of size $[\frac{|\underline{p}-\underline{p}_2|}{m}\sigma(\underline{p}', \underline{p}''; \underline{p}, \underline{p}_2)]\,dt$. It is therefore natural to call *collision volume* (per unit time) the quantity in square brackets: because it gives, after multiplication by the density of particles with momentum in $d\underline{p}$, the number of collisions per unit time and volume that particles with momentum $\underline{p}$ would undergo against a momentum $\underline{p}_2$ particle if there was only one such particle.

Introducing:

1.8.6
$$\begin{aligned}
f(\underline{p}', \underline{q})\, \mathrm{d}\underline{p}'\, \mathrm{d}q = {}& \text{number of particles with momentum } \underline{p}', \\
& \text{within } \mathrm{d}\underline{p}' \text{ in the cube } \mathrm{d}\underline{q} = \\
& \text{“number of collision centers”} \\
f(\underline{p}'', \underline{q})\, \mathrm{d}\underline{p}'' = {}& \text{density of particles with momentum } \underline{p}'', \\
& \text{within } \mathrm{d}\underline{p}'', \text{ in } \underline{q} = \\
& = \text{“density of particles that} \\
& \text{can undergo collision”} \\
\sigma(\underline{p}', \underline{p}''; \underline{p}, \underline{p}_2) = {}& \text{differential cross-section per unit solid} \\
& \text{angle for the considered collision,}
\end{aligned} \tag{1.8.6}$$

Note that the collision volume associated with a single collision center is, since the relative velocity at collision is $|\underline{p}' - \underline{p}''|/m = |\underline{p} - \underline{p}_2|/m$ n, also:

1.8.7
$$(|\underline{p}' - \underline{p}''|/m)\, \sigma(\underline{p}', \underline{p}'', \underline{p}, \underline{p}_2) \tag{1.8.7}$$

Hence the total number of collisions from Q', Q'' to Q_1, Q_2 is, per unit time:

1.8.8
$$f(\underline{p}'', \underline{q})\, \mathrm{d}\underline{p}''\, \frac{|\underline{p}' - \underline{p}''|}{m} \sigma(\underline{p}', \underline{p}'', \underline{p}, \underline{p}_2) f(\underline{p}', \underline{q})\, \mathrm{d}\underline{p}'\, \mathrm{d}\underline{q}\,, \tag{1.8.8}$$

clearly symmetric in $\underline{p}'$, $\underline{p}''$, although derived by treating $\underline{p}'$ and $\underline{p}''$ asymmetrically.

By a similar argument the number of "inverse" collisions is:

1.8.9
$$f(\underline{p}, \underline{q}) f(\underline{p}_2, \underline{q})\, \mathrm{d}\underline{p}\, \mathrm{d}\underline{p}_2\, \mathrm{d}\underline{q} \frac{|\underline{p} - \underline{p}_2|}{m}\, \sigma(\underline{p}, \underline{p}_2; \underline{p}', \underline{p}'')\,. \tag{1.8.9}$$

One then remarks that (1.8.5) imply:

1.8.10
$$\begin{aligned} &d\underline{p}'\,d\underline{p}'' = d\underline{p}\,d\underline{p}_2 &&(\text{“Liouville's theorem”})\\ &|\underline{p}'-\underline{p}''| = |\underline{p}-\underline{p}_2| &&(\text{“conservation of momentum”}); \end{aligned} \tag{1.8.10}$$

moreover the cross-section, as is in general true in collisions governed by central forces, depends exclusively on the *deflection angle* between $(\underline{p}'-\underline{p}'')$ and $(\underline{p}-\underline{p}_2)$ and on the relative speed $|\underline{p}'-\underline{p}''|/m$; and it is proportional to the (normalized) solid angle $d\omega$ into which $(\underline{p}-\underline{p}_2)$ points with respect to $(p'-p'')$.

Note in this respect that the collision final data, *i.e.* $\underline{p},\underline{p}_2$, do not determine $\underline{p}',\underline{p}''$ via (1.8.5) but they leave the direction $d\omega$ of $\underline{p}'-\underline{p}''$ arbitrary.

We shall then set $\sigma(\underline{p}',\underline{p}'',\underline{p},\underline{p}_2) = \sigma(\omega,|\underline{p}'-\underline{p}''|)\,d\omega = \sigma(\omega)\,d\omega$ where the last relation is only valid when the interaction between the spheres is assumed a *rigid sphere* interaction; and from the scattering theory it follows that in this case $\sigma(\omega)$ is independent of ω: $\sigma(\omega)=4\pi R^2$.

Hence (1.8.10) allow us to rewrite (1.8.8), (1.8.9) as:

1.8.11
$$\begin{aligned} &f(\underline{p}',\underline{q})f(\underline{p}'',\underline{q})\,d\underline{p}\,d\underline{p}_2\,d\underline{q}\,d\omega\,(|\underline{p}'-\underline{p}''|/m)\,\sigma(\omega)\\ &f(\underline{p},\underline{q})f(\underline{p}_2,\underline{q})\,d\underline{p}\,d\underline{p}_2\,d\underline{q}\,d\omega\,(|\underline{p}-\underline{p}_2|/m)\,\sigma(\omega) \end{aligned} \tag{1.8.11}$$

where, given $\underline{p},\underline{p}_2$, the vectors $\underline{p}'$, $\underline{p}''$ are computed from (1.8.5) and from the information that the solid angle between $\underline{p}-\underline{p}_2$ and $\underline{p}'-\underline{p}''$ is ω.

Introducing (1.8.11) in (1.8.4) and dividing by ε one finds the *Boltzmann equation*:

1.8.12
$$\begin{aligned} \frac{\partial f}{\partial t}(\underline{p},\underline{q})+\frac{\underline{p}}{m}\cdot\frac{\partial f}{\partial \underline{q}}(p,q) = \int \frac{|\underline{p}-\underline{p}_2|}{m}\,\sigma(\omega)\;d\omega\,d\underline{p}_2\cdot\\ \cdot\big(f(\underline{p}',\underline{q})f(\underline{p}'',\underline{q})-f(\underline{p},\underline{q})f(\underline{p}_2,\underline{q})\big) \end{aligned} \tag{1.8.12}$$

In (1.8.12) one supposes that $\underline{q}$ varies over the whole space: but the most interesting cases concern systems (one should say “rarefied gases” because of the conditions under which (1.8.12) has been derived) confined in a given volume V. In such cases (1.8.12) must be complemented by suitable *boundary conditions* that depend on the microscopic nature of the collisions of the particles against the walls.

Since the discussion of the boundary conditions is delicate we shall avoid it and in the case of confined systems we shall suppose, for simplicity, that periodic boundary conditions hold. This means that we imagine the volume V as a cube with opposite faces identified: *i.e.* a particle that collides with one of the cube faces reemerges, after the collision, from the opposite face and with the same velocity. For a deeper analysis of the problem of the boundary conditions (and in general of Boltzmann's equation) see [Ce69].

It should be clear that (1.8.12) is an approximation because we neglected:

(i) the possibility of multiple collisions,

(ii) the possibility that particles in the same volume element do not behave independently, as implicitly assumed in deriving (1.8.12), and instead, as time goes by, correlations between positions and velocities build up, and make certain collisions more probable than others, or multiple collisions relatively more probable with respect to binary ones. This approximation is sometimes called *molecular chaos*.

Such effects should disappear in the Grad-Boltzmann limit (1.8.3) provided they are absent at the initial time. This "conjecture" is known as *Grad's conjecture* on the validity of the "*Stosszahlansatz*", a word that, for traditional reasons, just denotes the lack of correlation between the motions of different particles at various instants of time.

In Appendix 1.A2 we show how an analysis of the corresponding conjecture can be easily performed in a much simpler case, in which a gas of particles moves in a space occupied by randomly placed spherical scatterers: a model called *Lorentz's model*. The particles collide with the scatterers but do not interact with each other, so that Boltzmann's equation for this model turns out to be linear.

Returning to Boltzmann's equation and postponing the analysis of the fundamental assumptions (i) and (ii) discussed above, the *irreversibility* of the approach to equilibrium that it implies can be demonstrated on the basis of the following remarks.

Multiplying both sides of (1.8.12) by $1, \underline{p}, \frac{1}{2m}\underline{p}^2$ or by $\log f(\underline{p},\underline{q})$ and integrating over $\underline{p}$ and $\underline{q}$ (under the assumption that $f(\underline{p},\underline{q}) \to 0$ rapidly as $(\underline{p},\underline{q}) \to \infty$ or, when $\underline{q}$ is restricted to a fixed container, that f satisfies suitable boundary conditions on the $\underline{q}$ coordinate) one finds that the quantities

$$\begin{aligned} N &= \int f(\underline{p},\underline{q})\,\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}, & \underline{P} &= \int \underline{p} f(\underline{p},\underline{q})\,\mathrm{d}\underline{p}\,\mathrm{d}\underline{q} \\ T &= \int (\underline{p}^2/2m) f(\underline{p},\underline{q})\,\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}\,, & H &= -\int f(\underline{p},\underline{q}) \log f(\underline{p},\underline{q})\,\mathrm{d}\underline{p}\,\mathrm{d}\underline{q} \end{aligned} \tag{1.8.13}$$

1.8.13

satisfy the relations:

1.8.14

$$\begin{aligned} &\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{\mathrm{d}\underline{P}}{\mathrm{d}t} = \frac{\mathrm{d}T}{\mathrm{d}t} = 0 \\ &\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{1}{4}\int \frac{|\underline{p}-\underline{p}_2|}{m}\sigma(\omega)\,\mathrm{d}\omega\big(f(\underline{p}',\underline{q})f(\underline{p}'',\underline{q}) - f(\underline{p},\underline{q})f(\underline{p}_2,\underline{q})\big)\cdot \\ &\qquad \cdot \big(\log f(\underline{p}',\underline{q})f(\underline{p}'',\underline{q}) - \log f(\underline{p},\underline{q})f(\underline{p}_2,\underline{q})\big)\,\mathrm{d}\underline{p}\,\mathrm{d}\underline{p}_2\,\mathrm{d}\underline{q} \,\ge\, 0 \end{aligned} \tag{1.8.14}$$

as can be checked by a simple calculation in which an essential role is played by the symmetry of the right-hand side of (1.8.12) between $\underline{p}, \underline{p}_2$ and $\underline{p}', \underline{p}''$ and $\mathrm{d}\underline{p}\,\mathrm{d}\underline{p}_2 = \mathrm{d}\underline{p}'\,\mathrm{d}\underline{p}''$ (*i.e.* Liouville's theorem, see (1.8.10), and momentum conservation, see (1.8.5))) is used together with the relations $\log x + \log y = \log xy$ and $(x-y)(\log x - \log y) \ge 0$.

Therefore while the first three relations in (1.8.14) imply five *conservation laws* (of the particle number, of momentum and of (kinetic) energy),

the fourth manifestly implies irreversibility and is called *Boltzmann's H-theorem*.

Furthermore (1.8.14) shows that the only possible equilibrium distributions $f(\underline{p},\underline{q})$ can be those for which

1.8.15
$$f(\underline{p}',\underline{q})f(\underline{p}'',\underline{q})=f(\underline{p},\underline{q})f(\underline{p}_2,\underline{q}) \tag{1.8.15}$$

where $\underline{p},\underline{p}_2,\underline{p}',\underline{p}''$ satisfy (1.8.5).

Equation (1.8.15) and the arbitrariness of $\underline{p},\underline{p}_2,\underline{p}',\underline{p}''$ imply, via a simple argument that we leave out,

1.8.16
$$f(\underline{p},\underline{q})=\frac{\rho(\underline{q})e^{-\beta(\underline{q})(\underline{p}-\underline{p}_0(q))^2/2m}}{(2\pi\beta(\underline{q})^{-1}m)^{3/2}} \tag{1.8.16}$$

where $\beta(\underline{q}),\underline{p}_0(\underline{q}),\rho(\underline{q})$ are arbitrary functions and the factor in the denominator of the right-hand side has been introduced for the purposes of a simple normalization, so that $\rho(\underline{q})$ could be interpreted as the density at the point $\underline{q}$ since it is then $\rho(\underline{q})=\int f(\underline{p},\underline{q})\,\mathrm{d}\underline{p}$.

This means that the equilibrium distributions $f(\underline{p},\underline{q})$ necessarily have the form (1.8.16). Considering subsequently the simple case of a system in a cubic container with periodic boundary conditions it is easy to show that, if f satisfies (1.8.12), (1.8.16) and $\partial f/\partial t=0$ (*i.e.* it is stationary) then it must *necessarily be* that $\beta(\underline{q}),\rho(\underline{q}),\underline{p}_0(\underline{q})$ are $\underline{q}$-independent.

In fact if f has the form (1.8.16) the right-hand side of (1.8.12) vanishes and, therefore, $\partial f/\partial t=0$ implies $\underline{p}\cdot\partial f/\partial\underline{q}=0$. Hence denoting by $\hat f(\underline{p},\underline{k})$ the Fourier transform of f with respect to q this implies that $\underline{p}\cdot\underline{k}\,\hat f(\underline{p},\underline{k})=0$, so that if $\hat f(\underline{p},\underline{k})$ is continuous in $\underline{p}$ it must be that $\hat f(\underline{p},\underline{k})=0$ for $\underline{k}\neq\underline{0}$. This means that f is $\underline{q}$-independent and that $\beta(\underline{q})$, $\rho(\underline{q})$, $\underline{p}_0(\underline{q})$ are constants.

We see that the H-theorem not only shows that the system evolves irreversibly, but it also shows that the one-particle distribution $f(\underline{p},\underline{q})$ evolves towards the free *Maxwell-Boltzmann distribution* which, one should not fail to note, is just a typical property of an element μ of the canonical ensemble in a system in which the interaction energy between the particles is so small (when their hard cores do not overlap) that the total energy of the system can be identified with the kinetic energy. The parameters $\beta,\rho,\underline{p}_0$ of this distribution are uniquely determined by the initial data via the conservation laws in (1.8.14).

It is natural to think that the H-theorem is, for rarefied gases, the microscopic version of the second law of thermodynamics which states that in isolated systems entropy increases (while equilibrium is approached): entropy should be identified as proportional to H.

It is therefore important to stress that *the H-theorem is manifestly in contrast with the reversibility properties of Newtons's equations* and, consequently, it cannot be a mathematical consequence of the latter, as already remarked, at least not in the literal, *i.e.* naive, sense of the word.

Just for this reason it becomes essential to understand whether this contrast between macroscopic irreversibility and microscopic reversibility can be overcome.

The alleged incompatibility between the two conflicting properties was the cause of violent critiques to Boltzmann who created the ergodic hypothesis (thus laying the foundations of the modern ergodic theory) in one of his attempts to answer his critics on a theoretical basis, solidly resting on mechanics rather than on the admittedly obscure Stosszahlansatz.

To investigate the question one can take two viewpoints, which appear somewhat to overlap in Boltzmann's brilliant and misunderstood (by his contemporaries) attempt at defending his theory and his H-theorem.

The first point of view is that the ergodic hypothesis holds, in the sense of (i) in §1.7 reinforced by (ii) and (iii), and therefore $\frac{dH}{dt} \ge 0$ could be only approximately true in the sense that it should hold "most of the time": when the cell $S^k\Delta$ that represents the microscopic state at the instant $k\tau$ runs through a great part of the ergodic cycle of given energy (*i.e.* the part in which the interesting macroscopic observables, see (1.5.5), are also constant for practical purposes). The relation $\mathrm{d}H/\,\mathrm{d}t \ge 0$ would then become false when $S^k\Delta$ exits such region.

The latter circumstance however can only happen, *in really macroscopic systems as well as in systems with few decades of particles*, with a temporal frequency longer, by far, than the longest astronomical scales, see §1.4, (1.4.3).

Therefore the system would for *all* practical purposes evolve irreversibly (and the evolution irreversibility would be symmetrical in time!). Reversibility could manifest itself over time scales *beyond eternity*, *i.e.* of many orders of magnitude greater than the age of the Universe, already for systems like a gas at normal conditions in a container of the size of a room, or of a very small box. Or, alternatively, for an *extremely short* time around the initial time: enough to "forget" the peculiarity of the preparation of the initial state.

A system set up initially in an "atypical" condition, *e.g.* occupying only half of the container, would expand to occupy the whole container and then it would continue to evolve without "ever" returning to occupy the initial half.

Of course if a *daemon* acting a few seconds after the initial time inverts all the velocities of *all* the particles of the system, then the system would retrace its previous evolution coming back to the initial state (but just for a very short time) and then it would evolve by again occupying the whole container proceeding towards equilibrium exactly as it would have done if its velocities had not been inverted (and, furthermore, according to an evolution law described approximately by Boltzmann's equation).

This inversion of the motion with production of an atypical situation after a short time (*i.e.* a non astronomically long time) from the initial instant requires the *exact* inversion of all velocities: if they were inverted with an error, even very small (provided not "astronomically small"), the system

would *not* go back and, instead, it would probably continue to evolve as if "nothing had happened" from a macroscopic point of view, [LV93]. The effort of the daemonic heavenly creature to intervene on earthly affairs, after leaving the realm of metaphysics, would therefore be in vain.

The second point of view has a more mathematical character and attempts at making quantitative the argument just described by relating it to Boltzmann's equation.

One imagines an initial datum in which particles, hard spheres with radius R, are independently distributed in phase space; we suppose that the density with which each of them is distributed is $\rho f_0(\underline{p},\underline{q})$ where f_0 is normalized to 1: $\int f_0(\underline{p},\underline{q})\,d\underline{p}\,d\underline{q}=1$.[8]

This system is evolved with Hamilton's equations (*i.e.* with elastic collision rules) and at time t one supposes that it is described by a distribution $\rho f_t(\underline{p},\underline{q})$, *without however assuming* that the particles are independently distributed; this means that the one-particle distribution $\rho f_t(\underline{p},\underline{q})$ provides only the information on the number of particles in $d\underline{p}\,d\underline{q}$ but *no longer* their correlations (as was the case at time 0, by construction) which will be non trivial, just because the "Stosszahlansatz" will not hold.

We now imagine, keeping t fixed, that $\rho\to\infty, R\to 0$ so that $\rho R^3\to 0$ but $\rho R^2=\lambda=$ fixed quantity: *i.e.* we consider the Grad-Boltzmann limit, see (1.8.3). If the above qualitative discussion is correct and if one remarks that in this limit the gas becomes a perfect gas (because the particles become point masses) in which equilibrium is attained by virtue of collisions between pairs of particles without two particles ever colliding more than once (because $R\to 0$ implies just this, as it is easy to estimate the probability of *recollision*, *i.e.* of the event in the Fig. 1.8.1.

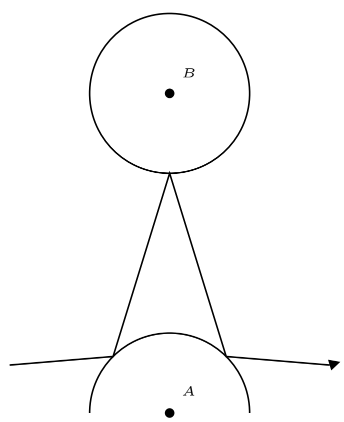


Fig. 1.8.1: Trajectory of C collides twice with A (A, B imagined fixed to simplify

[8] Note that if one throws randomly and independently hard spheres in a box then some of them may overlap. It is convenient *not* to exclude such a possibility provided one disregards completely the interaction of the overlapping spheres as long as they overlap and starts considering it only after they separate because of the motion: this is clearly a trick that introduces some minor simplification of the discussion while not affecting the macroscopic properties of the (rarefied) gas.

drawing).

as proportional to $\rho R^3 \to 0$, per unit time), then recalling that we denote ρf_t the evolving one particle density (to keep the integral of f_t normalized to 1) one has to conclude that the evolution of the limit $\lim_{R\to 0} f_t = \overline{f}_t$ should be *rigorously* described by the Boltzmann equation which for $\overline{f}_t$ is

1.8.17
$$\frac{\partial \overline{f}_t}{\partial t} - \frac{\underline{p}}{m} \cdot \frac{\partial \overline{f}_t}{\partial \underline{q}} = (\rho R^2) \int \frac{|\underline{p}' - \underline{p}''|}{m} \frac{\sigma(\omega)}{R^2} \, d\underline{p}_2 \, d\omega \cdot \\ \cdot \big(\overline{f}_t(\underline{p}', \underline{q}) \, \overline{f}_t(\underline{p}'', \underline{q}) - f_t(\underline{p}, \underline{q}) f(\underline{p}_2, \underline{q}) \big) \,. \tag{1.8.17}$$

One should note that this is ρ, R-independent because ρR^2 and $\sigma(\omega)/R^2$ are independent of R (recall that we are considering the case of hard sphere systems).

Hence the Boltzmann's equation should describe correctly the evolution of a rarefied gas for arbitrary times t: in fact we expect that in the Grad limit the recurrence times grow infinitely large while the collisions make the system *evolve on a time scale fixed by the flight time*: $((\rho R^2)\,\overline{v})^{-1}$. This can also be seen from (1.8.17) in which the time scale is fixed by $\rho R^2 |\underline{p}' - \underline{p}''|/m$ which, in the average, is just $\sim \rho R^2 \,\overline{v}$.

Equation (1.8.17) has been proved with complete mathematical rigor only recently and for times $t \ll 1/\rho R^2 \,\overline{v}$ for systems of hard spheres and for interesting classes of initial data f_0: this is the content of *Lanford's theorem* on the Grad conjecture, [La74].

This is an important confirmation, mathematically rigorous, of Boltzmann's point of view according to which reversibility, and the corresponding recurrence times, is not in contradiction with the experimental observation of irreversibility. Because the time scale over which reversibility manifests itself is *not observable* while that in which irreversibilty can be observed is related to the time of free flight $(\rho R^2 \,\overline{v})^{-1}$. Furthermore we see that irreversibility is not incompatible with the ergodic hypothesis, and Boltzmann's equation provides us with a model of the development of irreversible motions in situations in which the recurrence times are "infinitely" longer (even on astronomical scales) than the average time needed for a molecule to travel the free path (*i.e.* the flight time).

Thus Lanford's theorem, although it presents moderate interest for the applications due to the shortness of the interval of validity, $t \ll (\rho R^2 \,\overline{v})^{-1}$, has an enormous conceptual importance (apparently not yet fully appreciated by many) because it shows in a mathematically precise and rigorous fashion that there is *no incompatibility* between irreversible evolutions like the one described by the Boltzmann equation and the completely reversible Hamilton equations that describe the details of the microscopic motions. In fact mathematical rigor is particularly welcome here in consideration of the enormous amount of speculation on the theme and of pretended proofs of inconsistency between macroscopic irreversibility and mechanics. It has to

be hoped that, with time, Lanford's theorem will be appreciated as a basic advancement of statistical mechanics.

There is already a vast literature that developed following the spirit of Lanford's work (which was heralded by various works) and here I cannot discuss the matter further: for a proof developed with attention to the later developments and for the developments themselves the reader is referred to the recent treatise by Spohn, [Sp91], pp. 48-76.

This concludes our general introduction to statistical mechanics. We have seen that classical statistical mechanics holds only under certain conditions (see §1.2, for instance) at least as formulated here. It remains to analyze its consequences to deduce some of its applications and a better understanding of its validity and limitations.

Such an understanding is based, as already remarked, on the very consequences of the theory and it cannot be derived *a priori* as shown, for instance, by the fact that the basic condition in §1.2, namely $\vartheta_+/\vartheta_- > 1$ is compatible with very reasonable values of the temperature for "everyday physics" only because the intensity ε of the molecular interaction energy has order of magnitude of $\sim 10^{-14}\, erg$ and the radius of the molecules has size $\sim 2. \times 10^{-8} cm$. If this experimentally determined data had been very different the condition $\vartheta_+/\vartheta_- > 1$ could be impossible to satisfy at temperatures of importance for the observations usually carried out by classical thermodynamics. See Chap.II,III for a discussion of the latter points.

§1.9. A Historical Note. The Etymology of the Word "Ergodic" and the Heat Theorems

This section and Appendix 1.A1 are written in a way to be independent of the preceding sections: therefore there are here and there a few repetitions of subjects already analyzed in §1.1-§1.8. Few references to the previous sections are meant for readers familiar with them, but they are not essential for reading this section and Appendix 1.A1.

What follows is an expanded and revised version of various of my writings on Boltzmann's work, [Ga81], [Ga89].[9]

(1) *The etymology of the word "ergodic" and the heat theorems.*

Trying to find the meaning of the word *ergodic* one is led to a paper by Boltzmann, [Bo84]: see the footnote of S. Brush in his edition, [Bo64], of the *Lectures on Gas Theory*, on p. 297 (§5.10): here Boltzmann's paper is quoted as the first place where the word is introduced. Brush acutely warns the reader that Ehrenfests' paper misrepresents the opinions and even the terminology of Boltzmann and Maxwell and dates (in agreement with Gibbs,

[9] Readers might be interested in the referee report to one of my papers, [Ga95a], as it shows, in my opinion, how blind to evidence an historian of Science can be at times. The contents of the paper in question are reproduced here; the referee report and the corresponding unamended original version can be found in [Ga95b] (in English).

see p. vi of the introduction of [Gi81]), the first appearance of the concept to 1871, [Bo71b]. For instance the etymology that one finds in the Ehrenfests' paper is incorrect on this point: see [EE11], note #93, p.89, (where also the first appearance of the word is incorrectly dated and quoted).

In fact the basic idea of ergodicity can perhaps be traced to even earlier works, namely to the first work of Boltzmann on the theory of heat, [Bo66]: on p. 30 one finds that "*... this explanation is nothing else but the mathematical formulaton of the theorem according to which the paths that do not close themselves in any finite time can be regarded as closed in an infinite time"* (in this paper one also finds a general derivation of the necessity of the identification between average kinetic energy and absolute temperature).

The [Bo84] paper by Boltzmann is seldom quoted, I found only Brush's reference in [Bo64], and a partial account in [Br76], p. 242 and p. 368, before my own etymological discussion appeared in print in [Ga81], [Ga89], [Ga95a]. More recently the paper has been appropriately quoted by [VP92]; the paper was discussed also by [Ma88]. However no English translation of [Bo84] is available yet. Nevertheless I think that this is one of the most interesting papers of Boltzmann: it is a precursor of the work of Gibbs, [Gi81], on ensembles, containing it almost entirely (if one recalls that the equivalence of the canonical and microcanonical ensembles was already established (elsewhere) by Boltzmann himself, [Bo68], [Bo71]), and I will try to motivate this statement.

The paper stems from the important, not too well known, work of Helmholtz, [He95a], [He95b], who considered what we call today a system whose phase space contains only periodic orbits, or cycles of distinct energies: *i.e.* essentially a one-dimensional conservative system. He called such systems *monocyclic* systems and noted that they could be used to provide models of thermodynamics in a sense that Boltzmann undertakes to extend to a major generalization.

After an introduction, whose relative obscurity has been probably responsible for the little attention this paper has received, Boltzmann gives the notion of *stationary* probability distribution on the phase space of N interacting particles enclosed in a vessel with volume V. He calls a family $\mathcal{E}$ of such probabilities a *monode*, generalizing an "analogous" concept on monocyclic systems. In fact Boltzmann first calls a monode just a single stationary distribution regarded as an ensemble. But sometimes later he implicitly, or explicitly, thinks of a monode as a collection of stationary distributions parameterized by some parameters: the distinction is always clear from the context. Therefore, for simplicity, I here take the liberty of calling *monode* a collection of stationary distributions, and the individual elements of the collection will be called "elements of the monode".

The etymology that follows, however, is more appropriate for the elements of the monodes, as they are thought to consist of many copies of the same system in different configurations. By reading Boltzmann's analysis one can get the impression, see p. 132 of [Bo84], that the word "monode" had already been introduced by Maxwell, in [Ma79]; however the reference to

Maxwell is probably meant to refer to the notion of stationarity rather than to the word monode which does not seem to appear in [Ma79].

In fact the orbits of a monocyclic system can be regarded as endowed with a probability distribution giving an arc length a probability proportional to the time spent on it by the motion: hence their family forms a family of stationary probability distributions.

Etymologically, from the context of [Bo84], this appears to mean a family of stationary distributions with a "unique nature", (each consisting of systems with a "unique nature", differing only by the initial conditions), from μόνος and εἶδος , with a probable reference to Plato and Leibnitz. The concept appears, in fact, in some of Plato's dialogues, see the entry μονοειδής ("one in kind") in [LS94].

Then the following question is posed. Given an element μ of a monode $\mathcal{E}$ we can compute the average values of various observables, *e.g.* average kinetic energy, average total energy, average momentum transfer per unit time and unit surface in the collisions with the vessel walls, average volume occupied and density, denoted, respectively:

1.9.1 $$T = \frac{1}{N}\langle K\rangle_\mu, \quad U = \langle K+\Phi\rangle_\mu, \quad p, \quad V, \quad \rho = \frac{N}{V} \tag{1.9.1}$$

where Φ denotes the potential interaction energy and K the total kinetic energy. We then imagine varying μ in the monode $\mathcal{E}$, by an infinitesimal amount (this means changing any of the parameters which determine the element). *Question: is it true that the corresponding variations dU and dV are such that*:

1.9.2 $$\frac{dU + p\,dV}{T} \quad \text{is an exact differential} \quad dS\ ? \tag{1.9.2}$$

In other words is it true that the above quantities, defined in purely mechanical terms, satisfy the same relation that would hold between them if, for some thermodynamic system, they were the thermodynamic quantities bearing the same name, with the further identification of the average kinetic energy with the absolute temperature? (§1.5).

That the temperature should be identified with the average kinetic energy per particle was quite well established (for free gases) since the paper by Clausius, [Cl65], and the paper on the equipartition of kinetic energy by Boltzmann, [Bo66], [Bo68] (in the interacting cases); see the discussion of it in Maxwell's last scientific work, [Ma79]. The latter paper is also very interesting as Maxwell asks there whether there are other stationary distributions on the energy surface, and tries to answer the question by putting forward the ergodic hypothesis. If so the monode would provide a "mechanical model of thermodynamics" extending, by far, the early examples of Helmholtz on monocyclic systems.

Thus Boltzmann is led to the following definition, see §1.5, (1.5.6):

Definition: a monode $\mathcal{E}$ is called an orthode *if the property described by (1.9.2) holds.*

By reading [Bo84] the etymology of "orthode" is composed by ὀρθός and εἶδος, *i.e.* "right nature" or "correct nature".

The above deep definition has not been taken up by the subsequent literature. This is surprising, even more so as Boltzmann, in the same paper, proceeds to discuss "examples" of mechanical models of thermodynamics, *i.e.* examples of orthodic monodes. The above orthodicity concept is still attributed to Gibbs, see [Br76], p. 242.

The examples of orthodes discussed by Boltzmann in his paper are the *holode* and the *ergode* which are two ensembles whose elements are parameterized with two parameters β, N or U, N, respectively. Their elements are

1.9.3

$$\mu_{\beta,N}(d\underline{p}d\underline{q}) = \frac{d\underline{p}_1 \ldots d\underline{p}_n d\underline{q}_1 \ldots d\underline{q}_n}{const}\, e^{-\beta(K+\Phi)} \tag{1.9.3}$$

and

1.9.4

$$\mu_{U,N}(d\underline{p}d\underline{q}) = \frac{d\underline{p}_1 \ldots d\underline{p}_n d\underline{q}_1 \ldots d\underline{q}_n}{const}\, \delta(K(\underline{p}) + \Phi(\underline{q}) - U) \tag{1.9.4}$$

Boltzmann proves that the above two ensembles are both orthodes! thus establishing that the canonical and the microcanonical ensembles (using our modern terminology) are equilibrium ensembles and provide mechanical models of thermodynamics, see Chap.II for a discussion of similar proofs.

Boltzmann's simple proof makes use of the auxiliary (with respect to the above definition) notion of heat transfer. In the canonical case it yields exactly the desired result; in the microcanonical it is also very simple but somehow based on a different notion of heat transfer. An analysis of the matter easily shows, see §3.2 in Chap.II, that a definition of heat transfer for the microcanonical ensemble consistent with that of the canonical ensemble gives the result (1.9.2), *but only up to corrections* expected to be of order $O(N^{-1})$. As we have alluded to in §1.6 there is a problem only if one insists in defining in the same way the notion of heat transfer in the two cases: Boltzmann does not even mention this, possibly because he saw as obvious that the two notions would become equivalent in the thermodynamic limit.

Again from the context of [Bo84] one sees that the word "holode" has the etymological origin of ὅλος and εἶδος while "ergode" is a shorthand for "ergomonode" and it has the etymological root of ἔργον and εἶδος , meaning a "monode with given energy", [Ga81], [Ga95a].

The word "ergode" appears for the first time on p. 132 of [Bo84] but this must be a curious misprint, as the concept is really introduced on p. 134. On p. 132 the author probably meant to say "holode" instead; this has been correctly remarked by [VP92]. The above etymology was probably proposed for the first time by myself in various lectures in Roma, and it was included in the first section of [Ga81]. It has also been proposed in [Ja84], [Ma88]. The word "holode" is probably a shorthand for "holomonode", meaning a "global monode" (perhaps a monode involving states with arbitrary energy, *i.e.* spread over the whole phase space).

This is not what is usually believed to be the etymology of "ergode": the usual belief[10] comes from the Ehrenfests' statement that the etymology is ἔργον and ὁδός , with the meaning of "unique path on the surface of constant energy, see note #93 in [EE11]. The latter etymology has been taken up universally and has been attached to the subject of "ergodic theory", which is a theory dealing with time evolution properties.

(2) *The ergodic hypothesis, continuous and discrete phase space*

The etymological error of the Ehrenfests could be just an amusing fact: but it had a rather deep negative influence in the development of the 20-th century physics. They present their etymology in connection with the discussion (amounting to a *de facto* rejection) of the ergodic hypothesis of Boltzmann. In fact Boltzmann had come to the ergodic hypothesis in his attempts to justify *a priori* that the ergode, as a model of thermodynamics, had to produce *the* thermodynamics of a system with the given Hamiltonian function (and not just a model).

Boltzmann had argued that the trajectory of any initial datum evolves on the surface of constant energy, visiting all phase space points and spending equal fractions of time in regions of equal Liouville measure. See §1.3.

The Ehrenfests criticize such a viewpoint on surprisingly abstract mathematical grounds: basically they say that one can attach to each different trajectory a different label, say a real number, thus constructing a function on phase space which is constant on trajectories. Such a function would of course have to have the same value on points on the same trajectory (*i.e.* it would be a constant of motion). This is stated in the note #74, p. 86 where the number of different paths is even "counted", and referred to in the note #94, p. 89. Therefore, they conclude, it is impossible that there is a single path on the surface of constant energy, *i.e.* the ergodic hypothesis is inconsistent (except for monocyclic systems, for which it trivially holds).

The abstract mathematical nature of this argument, see also below for a critique, was apparently remarked on only by a mathematician, see [VP92] p. 86, (*i.e.* by Borel, 1914); but it escaped many physicists. It is worrying to note how literally so many took the Ehrenfests' version of the ergodic hypothesis and how easily they disposed of it, taking for granted that their formulation was the original one by Boltzmann and Maxwell, see [Br76], p. 383.

Having disposed of the ergodic hypothesis of Boltzmann, the Ehrenfests proceeded to formulate a new hypothesis, the rather obscure (and somewhat vague as no mention is made of the frequency of visits to regions in phase space) *quasi-ergodic hypothesis* see notes #98 and #99, p. 90, in [EE11]; it led physicists away from the subject and it inspired mathematicians to

[10] It is important, in this respect, to be aware that Boltzmann had studied the Greek language and, by his own account, quite well: see [Bo74], p. 133, to the point of having known at least small parts of Homer by heart. Hence there should be no doubt that he did distinguish the meanings of εἶδος and ὁδός which are among the most common words.

find the appropriate definition giving birth to ergodic theory and to its first nontrivial results.

The modern notion of ergodicity *is not* the quasi-ergodicity of the Ehrenfests. It is simply based on the remark that the Ehrenfests had defined a nontrivial constant of motion very abstractly, by using the axiom of choice. In fact from the definition, consisting in attaching a different number, or even $6N-2$ different numbers, to each distinct trajectory, there is *in principle* no way of constructing a table of the values of the function so defined in order to distinguish the different trajectories. In a system which is ergodic in the modern sense the Ehrenfests' construction would lead to a nonmeasurable function; and to a physicist endowed with common sense *such a function, which in principle cannot be tabulated, should appear as not existent, or as not interesting*. Thus motion on the energy surface is called ergodic if there are no *measurable* constants of motion: here measurable is a mathematical notion which essentially states the possibility of a tabulation of the function.

It is surprising that a generation of physicists could be influenced (in believing that the ergodic hypothesis of Boltzmann had to be abandoned as a too naive viewpoint) by an argument of such an exquisitely abstract nature, resting on the properties of a function that could not be tabulated (and not even defined if one did not accept the sinister axiom of choice). What is remarkable is the coincidence that the recognition and the development of the axiom of choice was due essentially to the same Zermelo who was one of the strongest opponents of Boltzmann's ideas on irreversibility; see also [Sc86].

Therefore it is worth, perhaps, trying to understand what Boltzmann may have meant when he formulated the ergodic hypothesis. Here one cannot fully rely on published work, as the question was never really directly addressed by Boltzmann in a critical fashion (he might have thought, rightly, that what he was saying was clear enough). The following analysis is an elaboration of [Ga81], [Ga95a] in some respects it gets quite close to [VP92]. It should be noted that [VP92] has a somewhat different point of view on several key issues, although we seem to share the main thesis that the [EE11] paper is responsible for most of the still persisting misunderstandings on Boltzmann's work, including the exclusive attribution to Gibbs of Boltzmann's ideas on ensembles, so clearly elaborated in [Bo84]. This is so even though, by reading the literature carefully, it is possible to realize that many were aware of the connection of Gibbs' work with Boltzmann's; see for instance [Br76], p. 242, first of all Gibbs, see p. vi of [Gi81] where he quotes the first section [Bo71c] of [Bo71b].

My point of view, adopted in the preceding sections, is that of those who believe that Boltzmann always conceived of the phase space and time as discrete spaces, divided into small cells, see [Bo72], p. 346. He always stressed that the continuum must be understood as a limit, see §1.1 (see also [Br76], p. 371, and [Kl62], [Kl72], [Kl73], [Du59]). The book by Dugas, [Du59], is particularly illuminating (also) on this respect (see for instance

Chap.I and the quotations of Boltzmann presented there, where he appears to identify the discrete viewpoint with the atomistic conceptions). In his writings Boltzmann very often makes this point: see for instance p. 42-44, note 4 on p. 51 (discrete time), p. 54, p. 168, p. 169, p. 243 (discrete time), p. 252/253, in [Bo74].

Although Boltzmann seems to have sometimes been quite apologetic about such a viewpoint (even calling it a "mathematical fiction", [Ba90], p.18, from [Bo72]; see also [VP92], p. 75), he took advantage of it to the point that one can say that most of his arguments are based on a discrete conception of phase space, followed at the end by a passage to the continuum limit, see §1.1. It should be understood however that the discretization that Boltzmann had in mind is by no means to be identified with the later concept of coarse graining; see Chap.IX where a modern version of Boltzmann's discretization is considered and where a distinction has to be made between cells and volume elements, see also [VP92] and [Ga95a].

It is easier for us, by now used to numerical simulations, to grasp the meaning of a "cell": in the numerical simulations a cell is simply an element of the discrete set of points in phase space, each represented within computer precision (which is finite). One should always discuss how much the apparently harmless discreteness of phase space affects results. This is, however, almost never attempted, see [Ga95a] for an attempt. A "volume element" in phase space has, instead, a size much larger than the machine resolution, so that it looks like a continuum (for some purposes). In the previous sections we have been careful to keep the discrete treatment of phase space always quite explicit, so that later we shall be easily able to see which are the consequences of a verbatim interpretation of the phase space discreteness.

Hence one can say that an essential characteristic of Boltzmann's thought is to have regarded a system of N atoms, or molecules, as described by a *cell* of dimension δq and δp in each position and momentum coordinates. He always proceeded by regarding such quantities as very small, avoiding entering into the analysis of their size, but every time this had some importance he seems to have regarded them as positive quantities.

A proof of this is when he refutes Zermelo's paradoxes by *counting* the number of cells of the energy surface of $1\,cm^3$ of normal air, [Bo96], a feat that can only be achieved if one considers phase space as discrete. His calculation has been discussed in §1.4, (1.4.3).

In particular this point of view must have been taken when he formulated the ergodic hypothesis: in fact conceiving the energy surface as discrete makes it possible to assume that the motion on it is "ergodic", *i.e.* it visits *all* the phase space points identified with cells, compatible with the given energy (and possibly with other "trivial" constants of motion), thus behaving as in a monocyclic system (as all the motions are necessarily periodic). This is in fact the definition in §1.3.

The passage to the continuum limit, which seems to have never been made by Boltzmann, of such an assumption is of course extremely delicate, and

it does not lead necessarily to the interpretation given by the Ehrenfests. It can easily lead to other interpretations, among which the modern notion of ergodicity, but it should not be attempted here, as Boltzmann himself did not attempt it.

In general one can hardly conceive that studying the continuum problem could lead to really new information, cannot be obtained by taking a discrete viewpoint. Of course some problems might still be easier if studied in the continuum, and the few results on ergodicity of physical systems do in fact rely explicitly on continuum models, [Si70]. However I rather interpret such results as illustrations of the complex nature of the discrete model: for instance the ergodicity theory of a system like billiards is very enlightening as it allows us to get some ideas on the question of whether there exist other ergodic distributions on the energy surface (in the sense of ergodic theory the answer is affirmative), and what is their meaning. The theory of the continuum models has been essential in providing new insights in the description of nonequilibrium phenomena, [RT71], [Ru78], [CELS93].

Finally the fruitfulness of the discrete models can be even more appreciated if one notes that they have been the origin of the quantum theory of radiation: it has even been maintained that Boltzmann had already obtained the Bose-Einstein statistics, [Ba90].

The latter is a somewhat strong intepretation of the 1877 paper, [Bo77]. The most attentive readers of Boltzmann have, in fact, noted that in his discretizations he uses, eventually, the continuum limit as a device to expedite the computations, manifestly not remarking that sticking to the discrete viewpoint would lead to *important differences* in some extreme cases. In fact he *does not discuss* the two main "errors", see Chap.III, that one commits in regarding a continuum formulation as an approximation (based on replacing integrals with sums), they were exploited for the first time by Planck, much later. The latter errors amount, in modern language, see Chap.III, to the identification of the Maxwell-Boltzmann statistics and the Bose-Einstein statistics, and to neglecting the variation of physically relevant quantities over the cells: see the lucid analysis in [Ku87], p.60; for a technical discussion see Chap.III.

The above "oversight" might simply be a proof that Boltzmann never took the discretization viewpoint to its extreme consequences, among which there is that the equilibrium ensembles are *no longer* orthodic in the sense of Boltzmann, see Chap.III, (although they still provide a model for thermodynamics provided the temperature is *no longer identified with the average kinetic energy*), a remark that very likely was not made by Boltzmann *in*

spite of his consideration and interest on the possibility of finding other integrating factors for the heat transfer dQ, see the footnote on p. 152 in [Bo84].[11]

The necessity of an understanding of this "oversight" has been in particular clearly advocated by Kuhn referring to Boltzmann's "little studied views about the relation between the continuum and the discrete", [Ku87], for instance.

There are many directions into which the analysis of the foundations of classical statistical mechanics can be developed. A somewhat different viewpoint for instance can be found in [Kr79]: this work of Krylov, and particularly part III has been very influential on Russian theoretical physics. In it, besides a very detailed critique of the foundations and of Boltzmann's and Gibbs' work, the foundations of the theory of the ergodicity of hard sphere systems is laid down: it was pursued later by Sinai. It also provided grounds for subsequent work on coarse graining (see Chap.IX) of Sinai and, in Sinai's interpretation, [Si79], also inspiration for the later theory of chaotic systems, [Si72], quite close to Ruelle's proposal, see Chap.IX and [Ru78c].

Appendix 1.A1. Monocyclic systems, Keplerian Motions and Ergodic Hypothesis

Consider a one-dimensional system with potential $\varphi(x)$ such that $|\varphi'(x)| > 0$ for $|x| > 0$, $\varphi''(0) > 0$ and $\varphi(x) \xrightarrow[x\to\infty]{} +\infty$. All motions are periodic so that the system is *monocyclic*. We suppose that the potential $\varphi(x)$ depends on a parameter V.

One defines *a state* a motion with given energy E and given V. And:

U = total energy of the system $\equiv K + \varphi$
T = time average of the kinetic energy K
V = the parameter on which φ is supposed to depend
p = $-$ average of $\partial_V \varphi$.

A state is parameterized by U, V and if such parameters change by dU, dV, respectively, we define:

1.*A*1.1
$$dL = -pdV, \qquad dQ = dU + pdV\,. \tag{1.A1.1}$$

Then:

Theorem (Helmoltz): *The differential* $(dU + pdV)/T$ *is exact.*

In fact let $x_\pm(U, V)$ be the extremes of the oscillations of the motion with given U, V and define S as:

[11] In checking my understanding of the original paper as partially discussed in [Ga81], I have profited from an English translation that Dr. J. Renn kindly provided me with later, (1984). He noticed this footnote in [Bo84] while performing his translation, (unfortunately still unpublished).

$$S=2\log\int_{x_-(U,V)}^{x_+(U,V)}\sqrt{K(x;U,V)}dx=2\log\int_{x_-(U,V)}^{x_+(U,V)}\sqrt{U-\varphi(x)}dx \tag{1.A1.2}$$

1.A1.2 so that

1.A1.3 $$dS=\frac{\int\big(dU-\partial_V\varphi(x)dV\big)\frac{dx}{\sqrt{K}}}{\int K\frac{dx}{\sqrt{K}}} \tag{1.A1.3}$$

and, noting that $\frac{dx}{\sqrt{K}}=\sqrt{\frac{2}{m}}dt$ we see that the time averages are given by integrating with respect to $\frac{dx}{\sqrt{K}}$ and dividing by the integral of $\frac{1}{\sqrt{K}}$. We find therefore

1.A1.4 $$dS=\frac{dU+pdV}{T} \tag{1.A1.4}$$

The above analysis admits an extension to keplerian motions: such systems are not monocyclic in the sense of Helmoltz, but if one considers only motions with a fixed eccentricity they have the same properties.

It is convenient to study motions in polar coordinates (ρ,ϑ), so that if $A=\rho^2\dot\vartheta, E=\frac{1}{2}m\dot{\underline{x}}^2-\frac{g\,m}{\rho}$, m being the mass and g the strength of the attraction due to gravity ($g=kM$ if k is the gravitational constant and M is the central mass) then

1.A1.5 $$E=\frac{1}{2}m\dot\rho^2+\frac{mA^2}{2\rho^2}-\frac{mg}{\rho}\,,\qquad \varphi(\rho)=-\frac{gm}{\rho} \tag{1.A1.5}$$

and

1.A1.6 $$\begin{aligned}&\dot\rho^2=\frac{2}{m}(E-\frac{mA^2}{2\rho^2}+\frac{mg}{\rho})\stackrel{def}{=}A^2(\frac{1}{\rho}-\frac{1}{\rho_+})(\frac{1}{\rho_-}-\frac{1}{\rho})\\ &\frac{1}{\rho_+\rho_-}=\frac{-2E}{mA^2},\quad \frac{1}{\rho_+}+\frac{1}{\rho_-}=\frac{2g}{A^2},\quad \frac{\rho_++\rho_-}{2}\stackrel{def}{=}a=\frac{mg}{-2E}\\ &\sqrt{\rho_+\rho_-}\stackrel{def}{=}a\sqrt{1-e^2},\qquad \sqrt{1-e^2}=\frac{A}{\sqrt{ag}}\,.\end{aligned} \tag{1.A1.6}$$

Furthermore if a motion with parameters (E,A,g) is periodic (hence $E<0$) and if $\langle\cdot\rangle$ denotes a time average over a period then

$$\begin{aligned}&E=-\frac{mg}{2a},\qquad \langle\varphi\rangle=-\frac{mg}{a},\qquad \langle\frac{1}{\rho^2}\rangle=\frac{1}{a^2\sqrt{1-e^2}}\\ &\langle K\rangle=\frac{mg}{2a}=-E,\qquad T=\frac{mg}{2a}\equiv\langle K\rangle,\qquad T_{ecc}=\big(1-\sqrt{1-e^2}\big)\,T\end{aligned} \tag{1.A1.7}$$

1.A1.7 Hence if S is defined by

1.A1.8 $$S=2\log\int_{\rho_-}^{\rho_+}\sqrt{\frac{2}{m}(E-\frac{mA^2}{2\rho^2}+\frac{mg}{\rho})}\,d\rho \tag{1.A1.8}$$

its differential is

1.A1.9
$$dS = \frac{dE - \frac{mA^2}{2a^2\sqrt{1-e^2}}\frac{dA^2}{A^2} + \frac{m}{a}dg}{\big(1-\sqrt{1-e^2}\big)\,T} \tag{1.A1.9}$$

This means that:

(1) If (E, A, g) are regarded as parameters then T_{ecc} is an integrating factor of:

1.A1.10
$$dQ = dE + p_A dA^2 + p_g dg, \qquad p_A = \frac{-m}{2a^2\sqrt{1-e^2}}, \quad p_g = \frac{m}{a} \tag{1.A1.10}$$

(2) Suppose that e is kept *constant*, so that the states as characterized by (E, g). Then using $\sqrt{1-e^2} = \sqrt{-2E/m}\,Ag^{-1}$ and $-E = T$, *i.e.*

1.A1.11
$$\frac{dA^2}{A^2} + \frac{d(-E)}{-E} - 2\frac{dg}{g} = 0 \tag{1.A1.11}$$

one can eliminate dA^2/A^2 from dS and find (after some simple algebra):

1.A1.12
$$dS = \frac{dE + (-2E)g^{-1}dg}{T} = d\log\frac{g^2}{-E} \tag{1.A1.12}$$

so that T is the integrating factor of $dQ = dE + pdV$ if $V = g$ and $p = \frac{-2E}{g} = \frac{m}{a}$, (Boltzmann). Note that the equations $pg = 2T$ and $E = -T$ can be interpreted as, respectively, analogues of the "equation of state" and the "ideal specific heat" laws (with the "volume" being g, the "gas constant" being $R = 2$ and the "specific heat" $C_V = 1$).

(3) If g is kept constant and (E, A^2) determine the states the integrating factor of $dQ = dE + p_A dA^2$, with $p_A = -\frac{m}{2a^2\sqrt{1-e^2}}$ is not the average kinetic energy T but the *eccentric temperature* T_{ecc}.

To check (2) note that by (1.A1.9), (1.A1.11)

1.A1.13
$$\begin{aligned} dS &= \frac{dE\,(1 - \frac{mA^2}{2a^2\sqrt{1-e^2}}\frac{1}{-E}) + dg\,(\frac{m}{a} - \frac{2mA^2}{2a^2\sqrt{1-e^2}}\frac{1}{g})}{\big(1-\sqrt{1-e^2}\big)\,T} = \\ &= \frac{(dE\, + \frac{m}{a}dg)(1 - \frac{A^2}{ag\sqrt{1-e^2}})}{\big(1-\sqrt{1-e^2}\big)\,T} = \frac{dE\, + \frac{m}{a}dg}{T} = \\ &= \frac{dE\, + \frac{-2E}{g}dg}{-E} = d\log\frac{g^2}{-2E}\,. \end{aligned} \tag{1.A1.13}$$

This concludes the discussion of Boltzman's version of Helmoltz's theory.

In general one can call a system *monocyclic* when it has the property that there is a curve $\ell \to \underline{x}(\ell)$, parameterized by its curvilinear abscissa ℓ, varying in an interval $0 < \ell < L(E)$, closed and such that $\underline{x}(\ell)$ covers all the positions compatible with the given energy E.

Let $\underline{x} = \underline{x}(\ell)$ be the parametric equations so that the conservation of energy can be written:

1.A1.14
$$\frac{1}{2}m \cdot \left(\frac{d\underline{x}}{d\ell}\right)^2 \dot{\ell}^2 + \varphi(\underline{x}(\ell)) = E\,. \tag{1.A1.14}$$

Then if we suppose that the potential energy φ depends on a parameter V and if T is the average kinetic energy, $p = -\langle \partial_V \varphi \rangle$ it follows:

1.A1.15
$$dS = \frac{dE + p\,dV}{T}, \qquad p = -\langle \partial_V \varphi \rangle, \quad T = \langle K \rangle\,. \tag{1.A1.15}$$

A typical case to which the above can be applied is the case in which the whole space of configurations is covered by the projection of a single periodic motion and the whole energy surface consists of just one periodic orbit, or at least only the phase space points that are on such an orbit are observable. Such systems provide natural models of thermodynamic behavior.

Noting that a chaotic system like a gas in a container of volume V will satisfy "for practical purposes" the above property we see that we should be able to find a quantity p such that $dE + pdV$ admits the inverse of the average kinetic energy as an integrating factor.

On the other hand the distribution generated on the surface of constant energy by the time averages over the trajectory should be an invariant distribution and therefore a natural candidate for it is the uniform distribution, *Liouville distribution*, on the surface of constant energy.

It follows that if μ is the Liouville distribution and T is the average kinetic energy with respect to μ then there should exist a function p such that T^{-1} is the integrating factor of $dE + p\,dV$.

Boltzmann showed that this is the case and, in fact, p is the average momentum transfer to the walls per unit time and unit surface, *i.e.* it is the *physical* pressure.

Clearly this is not a proof that the equilibria are described by the microcanonical ensemble. However it shows that for most systems, independently of the number of degrees of freedom, one can define a *mechanical model of thermodynamics*. The reason we observe approach to equilibrium over time scales far shorter than the recurrence times is due (as discussed in the previous sections) to the property that on most of the energy surface the actual values of the observables whose averages yield the pressure and temperature assume the same value. This implies that this value coincides with the average and therefore satisfies the *heat theorem*, as Boltzmann called the statement that $(dE + p\,dV)/T$ is an exact differential if p is the pressure (defined as the average momentum transfer to the walls per unit time and unit surface) and T is proportional to the average kinetic energy.

Appendix 1.A2. Grad-Boltzmann Limit and Lorentz's Gas

It is interesting to see how to derive Boltzmann's equation in simple models in which it becomes a linear equation. The models are well known since

Lorentz introduced them in his attempt to establish more firmly Drude's theory of electric conduction in metals.

In the models there are two types of particles: the W-particles (wind-particles) and the T-particles (tree-particles).

The W-particles move through space interacting only with the T-particles which, however, are supposed to be infinitely heavy compared to the W-particle and are supposed to be at rest and randomly distributed in space.

Each model is completely described by the $W-T$ interaction and by the T-particle distribution. From now on we shall focus our interest on the case in which the T-particles are distributed as the space distribution of a perfect gas (Poisson's distribution) with density n. We shall also assume that the T-particles are, *with respect to the W-particles*, hard spheres of radius a, reflecting the W-particles on their surface.

The assumed tree distribution is such that the probability for finding inside a given region Λ, with volume $V(\Lambda)$, exactly N tree particles, and for finding them in the infinitesimal cubes $d\underline{c}_1,\ldots,d\underline{c}_N$ around $\underline{c}_1,\ldots,\underline{c}_N$, is:

1.A2.1
$$f_\Lambda(\underline{c}_1,\ldots,\underline{c}_N)\frac{d\underline{c}_1,\ldots,d\underline{c}_N}{N!}=e^{-nV(\Lambda)}\frac{n^N}{N!}d\underline{c}_1,\ldots,d\underline{c}_N \tag{1.A2.1}$$

where the parameter n has the interpretation of density of the tree particles.

Note that, since the T-particles are hard spheres only with respect to the W-particles but not with respect to the each other, there are configurations $\underline{c}_1,\ldots,\underline{c}_N$ of trees in which the hard spheres overlap, (for some comments on this point see §1.6).

If $x=(\underline{p},\underline{q})$ is a W-particle phase space coordinate ($\underline{p}$= velocity, $\underline{q}$= position) the symbol

1.A2.2
$$S_t^{\underline{c}_1,\cdots\underline{c}_N}x \tag{1.A2.2}$$

will denote the W-particle coordinate $x'=(\underline{p}',\underline{q}')$ into which x evolves in time t in the presence of N tree-particles located at $\underline{c}_1,\ldots,\underline{c}_N$. The symbol $\omega(\underline{p})$ will denote the direction of $\underline{p}$ and $\hat{x}$ will denote the pair $(\omega(\underline{p}),\underline{q})$ if $x=(\underline{p},\underline{q})$.

Since the velocity $|\underline{p}|$ is conserved it is clear that $S_t^{\underline{c}_1,\ldots,\underline{c}_N}x$ depends only on the trees located within a distance $(|\underline{p}|t+a)$ from $\underline{q}$. The symbols:

1.A2.3
$$\left(S_t^{\underline{c}_1,\ldots,\underline{c}_N}x\right)_1\,,\ \left(S_t^{\underline{c}_1,\ldots,\underline{c}_N}x\right)_2\,,\ \omega\left(\left(S_t^{\underline{c}_1,\ldots,\underline{c}_N}x\right)\right) \tag{1.A2.3}$$

will, respectively, denote the velocity, position and momentum direction of (1.A2.2); and we also set:

1.A2.4
$$S_t^{\underline{c}_1,\ldots,\underline{c}_N}\hat{x}=\left(\omega\left(\left(S_t^{\underline{c}_1,\ldots,\underline{c}_N}x\right)\right)\,,\ \left(S_t^{\underline{c}_1,\ldots,\underline{c}_N}x\right)_2\right) \tag{1.A2.4}$$

Similarly we can give a natural meaning to the evolution of m W-particles:

1.A2.5
$$S_t^{\underline{c}_1,\ldots,\underline{c}_N}(x_1,\ldots,x_m)=\left(S_t^{\underline{c}_1,\ldots,\underline{c}_N}x_1,\ldots,S_t^{\underline{c}_1,\ldots,\underline{c}_N}x_m\right)\,. \tag{1.A2.5}$$

which takes into account the fact that there are no W-W interactions.

It is easy to derive Boltzmann's equation for W-particles in the case of the Lorentz' gas described above, see (1.A2.10) below. One realizes that the assumptions to be made in order to derive the Boltzmann's equation are essentially the same as conditions 1), 2), 3), of Sect. 8. They are:

(i) a W-particle never hits twice the same particle;

(ii) molecular chaos is assumed;

(iii) the size of the T-particles is negligible.

Here by a "chaotic" W-particle state we again mean a state such that the W-particle correlation functions are a product of one W-particle distribution: which are independent of the T-particle distribution. More precisely a chaotic state is such that the probability distribution for finding a certain configuration C of T-particles and a set of W-particles in $x_1,\ldots,x_m$ has the form $p(C)\prod_{i=1}^m f_0(x_i)$, where $p(C)$ denotes the distribution (1.A2.1) and this is interpreted as 0 if any wind particle is inside the hard cores of C.**[12]**

Clearly assumptions (i), (ii) and (iii) can be only approximately true.

Let us formulate Grad's limit conjecture for the Lorentz gas. Assume that the initial W-particles state is such that the probability density for finding W-particle in $dx_1\ldots dx_m$ is $m!^{-1}$ times:

1.A2.6
$$f(x_1,\ldots,x_m;\ 0)=\int_{C\ \mathrm{comp}\,(x_1,\ldots,x_m)} p(C)\prod_{i=1}^m f_0(x_i) \qquad (1.A2.6)$$

where $f_0(x)$ is a given function of x and the "integral" is the "sum" over all the T-particle configurations *compatible* with $x_1,\ldots,x_m$ (i.e. over the C's such that no W-particle is located inside the hard core of a T-particle). The *compatibility between* $(x_1,\ldots,x_m)$ *and* C *is expressed by the notation* $(x_1,\ldots,x_m)_{\mathrm{comp}}C$.

Note that (1.A2.6) is not a product state for the W-particles: this is so because here we have hard core interactions between them and the T particles.

Consider the state obtained by evolving the initial state (1.A2.6):

1.A2.7
$$f(x_1,\ldots,x_m;\ t)=\int_{C\ \mathrm{comp}\ (x_1,\ldots,x_m)} p(C)\prod_{i=0}^m f_0(S^C_{-t}x_j) \qquad (1.A2.7)$$

---

[12] Explicitly this means the following. Let p be the probability of finding the W-particles in a infinitesimal cube $dx_1\ldots dx_m$ around the configuration $X=(x_1,\ldots,x_m)$ in the box Λ_0, and a tree configuration in the infinitesimal cube $d\underline{c}_1\ldots d\underline{c}_M$ around $C=(\underline{c}_1,\ldots,\underline{c}_M)$ in the box Λ, assuming it wider by an amount a than Λ_0, at least. Here $x_i=(\underline{p}_i,\underline{q}_i)$. Then p is the product of the probability in (1.A2.1) times $m!^{-1}(\prod_{i=1}^m f_0(x_i)\,dx_i)\,e^{-\int_C f_0(\underline{\xi})d\underline{\xi}}$, where $\underline{\xi}=(\underline{p},\underline{q})$ and $\int_C d\underline{\xi}$ means integration over $\underline{p}$ and over the $\underline{q}\in\Lambda_0$ which are outside the hard spheres centered on $C=(\underline{c}_1,\ldots,\underline{c}_N)$. In other words the W particles also have a Poisson distribution, in the region outside the T particles, with a density function f_0.

and then let the T-particle density n tend to infinity and the hard core W-T radius tend to zero in such a way that $na^3 \to 0$ but $na^2 \to l \neq 0, \infty$. We shall imagine that the solid angle integration is normalized so that $\int d\underline{\omega} = 1$. Grad's conjecture can then be formulated as:

If $t \geq 0$ and under "mild assumptions" on f_0, the following limit exists:

1.A2.8
$$\lim_{\substack{na^3 \to 0 \\ na^2 \to \text{ const } \neq 0,\infty}} f(x_1, \ldots, x_m;\ t) = \tilde{f}(x_1, \ldots, x_m;\ t) \tag{1.A2.8}$$

and:

1.A2.9
$$\tilde{f}(x_1, \ldots, x_m;\ t) = \prod_{i=1}^{m} \tilde{f}(x_i;\ t) \tag{1.A2.9}$$

and $\tilde{f}(x;t)$ satisfies the Boltzmann equation:

1.A2.10
$$\frac{\partial \tilde{f}}{\partial t}(x,t) + \underline{p} \cdot \frac{\partial \tilde{f}}{\partial \underline{q}}(x,t) = \lambda^{-1}|\underline{p}| \int (\tilde{f}(x',t) - \tilde{f}(x,t))\sigma(\omega)\, d\omega \tag{1.A2.10}$$

where $x = (\underline{p}, \underline{q}), x' = (\underline{p}', \underline{q})$ and $\underline{p}'$ is a vector with the same length as $\underline{p}$ but forming an angle ω with it; $a^2\sigma(\underline{\omega}) = a^2$ is the scattering cross-section of a hard sphere with radius a and $\lambda^{-1} = 4\pi na^2$.

A similar conjecture can be formulated in a two-dimensional model; here the solid angle ω has to be replaced by the deflection angle β (see Fig. 1.A2.2) and $\sigma(\omega)$ by $\sigma(\beta) = \frac{\pi}{2}\sin\frac{\beta}{2}$ and $\lambda^{-1} = 2an$. Of course the Boltzmann limit will be, in this case, $na^2 \to 0$, $2na \to \lambda^{-1} \neq 0, \infty$.

It is easy to construct a proof of the above conjecture in the two-dimensional case. The three-dimensional case could be treated along the same lines as will become apparent from the proofs. We shall assume, for simplicity, the spatial dimension to be two. The direction $\omega(\underline{p})$ will be in this case the angle ϑ between $\underline{p}$ and a fixed axis.

We first specify the "mild assumptions" on f_0. The function $f_0(x)$ will be thought as $f_0(|\underline{p}|, \omega(\underline{p}), \underline{q})$, if $x = (\underline{p}, \underline{q})$, and we can write:

1.A2.11
$$f_0(|\underline{p}|, \omega(\underline{p}), \underline{q}) = \int d\,\underline{q}'\, d\omega'\, f_0(|\underline{p}|, \omega', \underline{q}')\delta(\underline{q} - \underline{q}')\delta(\omega(\underline{p}) - \omega') \tag{1.A2.11}$$

we shall abbreviate (ω', q') to ξ, $d\underline{q}'d\omega'$ to $d\xi$ and $\delta(\underline{q} - q')\delta(\omega(\underline{p}) - \omega')$ to $\delta(x - \xi)$. Hence, by using definition (1.A2.4), Eq. (1.A2.7) becomes, for $m = 1$:

1.A2.12
$$f(x;t) = \int d\,\xi f_0(|\underline{p}|, \xi) \int_{C \text{ comp } x} \delta(S^C_{-t}\hat{x} - \xi)p(C)\,. \tag{1.A2.12}$$

It is therefore useful to consider the function:

1.A2.13
$$g(\xi; x; t) = e^{\pi na^2} \int_{C \text{ comp } x} p(C)\delta(S^C_{-t}\hat{x} - \xi) \tag{1.A2.13}$$

where the factor $e^{\pi n a^2}$ has been introduced for normalization purposes (note that it tends to 1, in the Boltzmann limit). It is easily checked that:

1.A2.14
$$g(\xi;x;0) = \delta(\hat{x}-\xi), \qquad \int g(\xi;x;t)d\xi \equiv 1$$
$$f(x,t) = e^{-\pi n a^2} \int d\,\xi\, f_0(|\underline{p}|,\xi)\, g(\xi;x;t)\,. \tag{1.A2.14}$$

We shall show that as $na^2 \to 0$, $2na \to \lambda^{-1} \neq 0, \infty$ the function $g(\xi;x;t)$ will tend to a limit $\tilde{g}(\xi;x;t)$ which satisfies the two dimensional analogue of relation (1.A2.10) with initial condition $\tilde{g}(\xi;x;0) = \delta(\hat{x}-\xi)$ and $|\underline{p}|$ fixed. The linearity of (1.A2.10), and of the third (1.A2.14), will imply, under suitable assumptions on f_0, that also $\tilde{f}(x,t)$ satisfies (1.A2.10).

We will not insist in discussing in which sense $g(\xi;x;t)$ converges to $\tilde{g}(\xi;x;t)$. It will appear from the proofs below that, if $x(t) = (\underline{q} + \underline{p}t, \underline{p})$, at least $g_0(\xi;x;t) \stackrel{def}{=} g(\xi;x;t) - e^{-\lambda|\underline{p}|t}\delta(\hat{x}-\hat{x}(t))$ converges to $\tilde{g}_0(\xi;x;t) \stackrel{def}{=} \tilde{g}(\xi;x;t) - e^{-\lambda|\underline{p}|t}\delta(\hat{x}-\hat{x}(t))$ pointwise for $t \neq 0$, and in the sense of the distributions for all $t \geq 0$. However a close examination of the proof will provide evidence against any uniformity of the convergence in t, unless t is restricted to a bounded interval (for further remarks on this point see below).

Under the above convergence conditions, "mild assumptions" could, for instance, be continuity and boundedness of f_0. The proof is based on a simple change of variables in (1.A2.13).

Let $x = (\underline{p}, \underline{q})$ and let $R(x,t)$ be the sphere with center $\underline{q}$ and radius $(|\underline{p}|t + a)$; then $S^{c}_{-t}x$ depends only on the T-particles in c contained in $R(x,t)$. Hence the integral (1.A2.13) can be explicitly written as:

1.A2.15
$$\begin{aligned} &g(\xi;x;t) = \\ &= e^{\pi n a^2} \sum_{M=0}^{\infty} \int_{R(x,t)^M} e^{-nV(R(x,t))} \frac{n^M}{M!} \delta\left(S^{\underline{c}_1,\ldots,\underline{c}_M}_{-t}\hat{x} - \xi\right) d\underline{c}_1 \ldots d\underline{c}_M \end{aligned} \tag{1.A2.15}$$

where $V(R(x,t))$ = area of $R(x,t)$ and where use has been made of the assumed Poisson distribution of the T-particles (1.A2.1).

Note that, in general, not all the T-particles $\underline{c}_1, \ldots, \underline{c}_M$ in (1.A2.15) will be hit by the trajectory $S^{\underline{c}_1,\ldots,\underline{c}_M}_{-\tau}x$, $0 \leq \tau \leq t$. Let $A_{x,t,N}$ denote the set of configurations $\underline{c}_1, \ldots, \underline{c}_N$ of N T-particles such that a W-particle with initial coordinate x hits, in the time t, all the N particles in $\underline{c}_1, \ldots \underline{c}_N$ at least once. We deduce from (1.A2.15), see Fig. 1.A2.1:

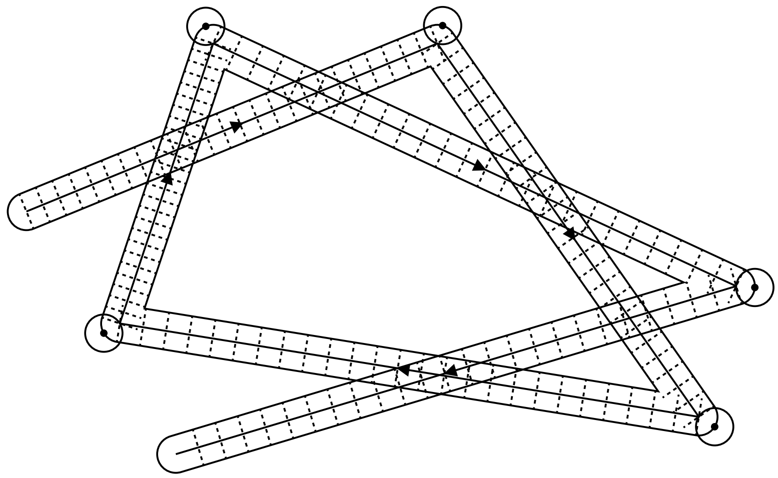


Fig. 1.A2.1: The set $P(t;\underline{c}_1,\ldots,\underline{c}_N)$ is the dashed region. The circles represent trees $\underline{c}_1,\ldots,\underline{c}_N$, ($N=5$) and the length of the trajectory in the dashed region is $|\underline{p}|t$.

1.A2.16
$$g(\xi;x;t) = e^{\pi n a^2}\sum_{N=0}^{\infty}\int_{A_{x,t,N}} n^N \tfrac{d\underline{c}_1,\ldots,d\underline{c}_N}{N!}\,\chi_{\underline{c}_1,\ldots,\underline{c}_N}(x)\cdot\delta\left(S_{-t}^{\underline{c}_1,\cdots\underline{c}_N}\hat{x}-\xi\right)\cdot \cdot\left[\sum_{M=N}^{\infty}\int_{\mathcal{R}}\tfrac{d\underline{c}'_1,\ldots,d\underline{c}'_{M-N}}{(M-N)!}\,n^{M-N}e^{-nV(R(x,t))}\right] \qquad (1.A2.16)$$

$\mathcal{R}\equiv$ set of points in $R(x,t)^{M-N}$ such that$\underline{c}'_1,\ldots,\underline{c}'_{M-N}\notin P(t;\underline{c}_1,\ldots,\underline{c}_N)$

where $\chi_{\underline{c}_1,\ldots,\underline{c}_N}(x)$ is 1 if x is compatible with the hard cores of $\underline{c}_1,\ldots,\underline{c}_N$ and 0 otherwise: the region $P(t;\underline{c}_1,\ldots,\underline{c}_N)$ is the tube like region (see Fig. 1.A2.1) swept by an ideal T-particle when its center is moved along the path $S_{-\tau}^{\underline{c}_1,\ldots,\underline{c}_N}x$, $0\le\tau\le t$.
The sum within square brackets in (1.A2.16) can be performed (since the integrals are trivial) and yields:

1.A2.17
$$e^{-nV(P(t;\underline{c}_1,\ldots,\underline{c}_N))} \qquad (1.A2.17)$$

so that $g(\xi,x;t)$ is:

1.A2.18
$$g(\xi;x;t) = e^{n\pi a^2}\sum_{N=0}^{\infty}\int_{A_{x,t,N}} n^N e^{-nV(P(t;\underline{c}_1,\ldots,\underline{c}_N))}\cdot\chi_{\underline{c}_1,\ldots,\underline{c}_N}(x)\cdot \cdot\delta\left(S_{-t}^{\underline{c}_1,\ldots,\underline{c}_N}\hat{x}-\xi\right)\frac{d\underline{c}_1\ldots d\underline{c}_N}{N!} \qquad (1.A2.18)$$

The reader should note the very simple probabilistic meaning of this equation which makes it almost self-evident: the T-particles in $A_{x,t,N}$ can be hit more than once in the time t. Divide $A_{x,t,N}$ as $A^1_{x,t,N}\cup A'_{x,t,N}$ where $A^1_{x,t,N}$ is the set of T-configurations in $A_{x,t,N}$ such that all their T-particles are hit just once by the trajectory $S_{-\tau}^{\underline{c}_1,\ldots,\underline{c}_N}x$, $0\le\tau\le t$.

To this decomposition of $A^1_{x,t,N}$ there corresponds a decomposition $g(\xi;x;t) = g_1(\xi;x;t) + g'(\xi;x;t)$ with

1.A2.19
$$g_1(\xi;x;t) = e^{\pi n a^2} \sum_{N=0}^{\infty} \int_{A^1_{x,t,N}} n^N \frac{d\underline{c}_1,\ldots,d\underline{c}_N}{N!} \chi_{\underline{c}_1,\ldots,\underline{c}_N}(x) \cdot \delta\left(S_{-t}^{\underline{c}_1,\ldots,\underline{c}_N} \hat{x} - \xi\right) e^{-nV(P(t;\underline{c}_1,\ldots,\underline{c}_N))} \tag{1.A2.19}$$

We now perform the change of variables, illustrated in Fig. 1.A2.2 from the $2N$ variables $\underline{c}_1,\ldots,\underline{c}_N$ to the new $2N+1$ variables $l_1,\ldots,l_{N+1}$, $\beta_1,\ldots,\beta_N$; we get

1.A2.20
$$\frac{d\underline{c}_1,\ldots,d\underline{c}_N}{N!} = a^N \delta\left(\sum_{i=1}^{N+1} l_i - |\underline{p}|t\right)\left(\prod_{i=1}^{N+1} dl_j\right) \prod_{j=1}^{N}\left(\frac{d\beta_j}{2}\,\sin\frac{\beta_j}{2}\right) \tag{1.A2.20}$$

represented as:

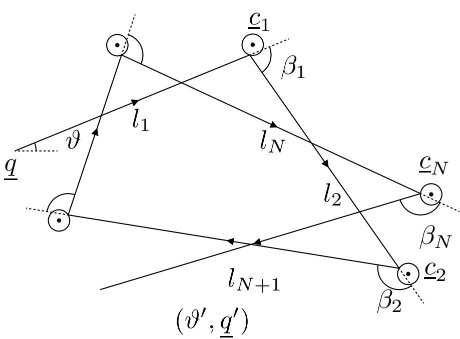


Fig. 1.A2.2

Hence the N-th order contribution to (1.A2.19) is given by (if $\omega(\underline{p}) = \vartheta$ and $x = (p,q) = (|p|,\vartheta,\underline{q})$, $\xi = (|p|,\vartheta',q')$):

1.A2.21
$$* \; e^{\pi n a^2}(2na)^N \int_0^\infty \prod_{i=1}^{N+1} dl_i \int_0^{2\pi} \prod_{i=1}^{N} (\sin\frac{\beta_i}{2}\,\frac{d\beta_i}{4}) \delta(\sum_{i=1}^{N+1} l_i - |\underline{p}|t) \cdot \cdot\, \delta(\sum_{i=1}^{N+1} \underline{l}_i - (\underline{q}' - \underline{q}))\delta(\sum_{i=1}^{N} \beta_i - (\vartheta' - \omega(\underline{p})))e^{-nV(P(t;\underline{c}_1,\ldots,\underline{c}_N))} \tag{1.A2.21}$$

where $\underline{l}_i$ are the vectors represented by arrows in Fig. 1.A2.2 ($|\underline{l}_1| = l_i$); the * in (1.A2.21) means that there is an extra condition on the integration region. It is the condition that none of the spheres of radius a around $\underline{c}_1,\ldots,\underline{c}_N$ has intersection with the straight segments of the broken line representing the trajectory in Fig. 1.A2.2 (i.e. this is the condition that $\underline{c}_1,\ldots,\underline{c}_N$ really belongs to $A^1_{x,t,N}$). Of course in (1.A2.21), $\delta\left(\sum_{i\equiv 1}^{N} \beta_i - (\vartheta' - \omega(\underline{p}))\right)$ means $\sum_{h=-\infty}^{+\infty} \delta(\sum_i \beta_i - (\vartheta' - \omega(\underline{p})) - 2\pi h)$.

In the limit $na^2 \to 0$, $2na \to \lambda^{-1} \neq 0, \infty$ the restrictions indicated by the * in (1.A2.21) become unimportant and $nV(P(t; \underline{c}_1, \ldots, \underline{c}_N))$ simplifies enormously:

1.A2.22
$$nV(P(t; \underline{c}_1, \ldots, \underline{c}_N)) \to 2na \sum_{j=1}^{N+1} l_j = \lambda^{-1}|p|t \tag{1.A2.22}$$

Hence the limit $\tilde{g}(\xi; x; t)$ as $na^2 \to 0$ and $2na \to \lambda^{-1} \neq 0, \infty$ of $g_1(\xi; x; t)$ is:

1.A2.23
$$\sum_{N=0}^{\infty} \lambda^{-N} \int_0^{\infty} \int_0^{2\pi} \Big(\prod_{i=1}^{N} \sin\frac{\beta_j}{2} \frac{d\beta_i\, dl_j}{4}\Big) \delta(\sum_{i=1}^{N+1} l_i - |\underline{p}|t) \cdot \\ \cdot\, \delta\left(\sum_{i=1}^{N+1} \underline{l}_i - (\underline{q}' - \underline{q})\right) \cdot \delta\left(\sum_{i=1}^{N} \beta_i - (\vartheta' - \omega(\underline{p}))\right) \cdot e^{-\lambda^{-1}|p|t} \tag{1.A2.23}$$

In the derivation of (1.A2.23) we have systematically disregarded convergence problems connected with the summation over N, M, *etc.*, since they are trivial as a consequence of the presence of the factorials and of the boundedness of the integration regions. The limit (1.A2.23) is pointwise for $t \neq 0$ and it could be checked that it holds also in the sense of distributions for $t \geq 0$.

Furthermore it could be checked that for $t > 0$ the function $g(\xi, x, t) \geq g_1(\xi, x, t)$ is bounded above by a $L_1(d\xi)$ function once the delta function contribution coming from the collisionless paths is subtracted to both terms; hence the limit (1.A2.23) holds also in the $L_1(d\xi)$ sense. Finally, by direct computation, it follows from (1.A2.23) that:

1.A2.24
$$\int \tilde{g}(\xi; x; t) d\xi \equiv 1 \tag{1.A2.24}$$

and this fact, together with the above convergence properties and (1.A2.14), implies the validity of the limit relation: $\lim_{\substack{na^2 \to 0 \\ 2na \to \lambda^{-1}}} g(\xi; x; t) - e^{-\lambda|\underline{p}|t} \delta(\hat{x} - \hat{x}(t)) = \tilde{g}(\xi; x; t) e^{-\lambda|\underline{p}|t} \delta(\hat{x} - \hat{x}(t))$ in $L_1(d\xi)$ for $t > 0$; furthermore it could be proved that this limit holds, for $t \geq 0$, in the sense of the distributions.

That (1.A2.23) is a solution of the Boltzmann equation can be checked directly by substituting $\tilde{g}$ into (1.A2.10), with initial condition $\tilde{g}(\xi; x; 0) = \delta(\hat{x} - \xi)$ and $|p|$ fixed, see for instance [Ga69] or check directly (recalling that $\sigma(\omega) d\omega = \frac{\pi}{2} \sin\frac{\beta}{2} \frac{d\beta}{2\pi}$ as, with our conventions, $d\omega = \frac{d\beta}{3\pi}$ if ω is the "solid angle" in the direction β.

To complete the proof of Grad's limit conjecture it remains to deal with the m-particle distributions. However we skip this point since it involves straightforward calculations based on changes of variable of the type illustrated in Fig. 1.A2.2.

We have thus described a proof of the Boltzmann limit conjecture in the case of a two-dimensional Lorentz gas with hard core W-T interactions

and free gas distribution of the T-particles. The generalization to three dimensions would be trivial. See also [LS82] where a more detailed and careful study of the mathematical aspects of the above analysis is performed with further insights and applications.

A less trivial generalization would be obtained by keeping the hard core W-T interaction but assuming that the T-particles are spatially distributed as if they were a gas of hard spheres with hard core size being proportional to the W-T radius. Other generalizations are conceivable in the direction of allowing soft $W - T$ particle interactions and more general T-particle distributions.

Much more difficult and interesting would be the treatment of *Knudsen's model*, in which the T-particles are allowed to move without suffering changes in their momentum in the collisions with the W-particles.

Had we done the calculations associated with the proof of (1.A2.9), we would have also found evidence of a lack of uniformity of the Boltzmann limit in the number m of W-particles even at fixed t: the larger m is, the closer one has to get near the Boltzmann limit in order to see factorization of the W-particle correlations.

We also wish to remark that even when the Boltzmann limit conjecture is true, one cannot expect that the solution $\rho\,\overline{f}(\underline{r},\underline{v},t)$ to the Boltzmann's equation (see §1.8 and (1.A2.8) above) is such that $\overline{f}(\underline{r},\underline{v},t)$ is a good approximation to the actual distribution $f(\underline{r},\underline{v},t)$ for large t: in fact one intuitively expects that for times of the order of $t_{m.f.p.}/na^3$ some nontrivial correlations will start building up thus destroying the molecular chaos and spoiling the validity of the Boltzmann equation.

This last remark is quite deceiving since it tells us that we cannot use, *without further assumptions*, the Boltzmann equation to investigate the long time behavior and, in particular, to compute the transport coefficients. From a rigorous point of view we cannot even be sure that the lowest order in na of the transport coefficients is correctly given by the value obtained in the Boltzmann limit. However it seems reasonable that this is, indeed, the case at least if the dimension of the space is larger than two (in one dimension a simple counterexample can be found by using soluble models [LP66]; in this case, however, the Boltzmann equation is *a priori* not expected to be a good approximation).

For further reading on the Lorentz gas see [WL69], [LS82].

The idea of the Boltzmann limit is clearly stated in [Gr58], see p. 214; the present proof in the case of the Lorentz gas is done in [Ga69] (for the case of $\tilde{g}(\underline{\xi};x;t)$ only) and was inspired by discussions and suggestions from J.L. Lebowitz.

Chapter *II*:

Statistical Ensembles

.

§2.1. Statistical Ensembles as Models of Thermodynamics

Given a mechanical system its microscopic states are described by the microscopic configurations of N (identical for simplicity) particles with mass m wandering in a given volume V: such configurations are represented by phase space cells of equal phase space volume h^{3N}.

The cells have dimensions δp and δq in momentum and position coordinates and they represent the maximal resolution with which we suppose that the microscopic states can possibly be observed: since we suppose the particles to be identical the phase space cells differing by a permutation of the particles must be regarded as identical. The parameter $h = \delta p \delta q$ empirically represents the precision with which the microscopic states can be determined, see Chap.I, §1.1 and §1.2.

Time evolution transforms cells into cells in a small time τ: so that cell Δ is transformed into $\Delta' = S\Delta$ by a transformation S defined in terms of the *total energy* or *Hamiltonian* function $E(\Delta)$, the sum of the kinetic energy $K(\underline{p})$ and the total potential energy $\Phi(\underline{q})$:

2.1.1
$$E(\Delta) = E(\underline{p},\underline{q}) = K(\underline{p}) + \Phi(\underline{q}) = \sum_{i=1}^{N} p^2/2m + \sum_{i<j} \varphi(\underline{q}_i - \underline{q}_j) \quad (2.1.1)$$
$$E(\underline{p},\underline{q}) \ge U_{\min} = \min E(\underline{p},\underline{q}) > -\infty$$

where $\underline{p} = (\underline{p}_1, ..., \underline{p}_N)$, $\underline{q} = (\underline{q}_1, ..., \underline{q}_N)$ are the momentum and position coordinates of the N particles and φ is interaction potential between particles, see §1.2. The second of (2.1.1) is a *stability* constraint that we shall assume to hold for all N (with $U_{\min}$ dependent on N): without it many of the integrals that we shall consider would be divergent.

In fact we shall see that the properly significant physical condition is that $U_{\min}$ can be taken $\ge -BN$ for some B; see (2.2.17) below.

We have then considered the stationary probability distributions μ that associate with every cell, *i.e.* with every microscopic state, its probability $\mu(\Delta)$ so that $\mu(\Delta) = \mu(S\Delta)$.

Families $\mathcal{E}$ of stationary distributions can be identified with families of macroscopic equilibrium states in which a generic observable f, *i.e.* a generic function defined on the phase space cells, takes an average value in the state $\mu \in \mathcal{E}$:

2.1.2
$$\overline{f} = \sum_{\Delta} \mu(\Delta) f(\Delta) \,. \quad (2.1.2)$$

Given a family $\mathcal{E}$ of stationary distributions on the space of microscopic states one can consider the averages that the most physically relevant observables take in a state $\mu \in \mathcal{E}$:

$$U(\mu) = \sum_{\Delta} \mu(\Delta) E(\Delta) \qquad \text{“energy”}$$

2.1.3
$$\begin{aligned} V(\mu) &= V(\Delta) = V && \text{“volume”} \\ K(\mu) &= \sum_{\Delta} \mu(\Delta) K(\Delta) && \text{“kinetic energy”} \\ P(\mu) &= \sum_{\Delta} \mu(\Delta) P(\Delta) && \text{“pressure”} \end{aligned} \tag{2.1.3}$$

where $P(\Delta)$ is the momentum variation per unit time and per unit surface area undergone by the particles in the microscopic state Δ in the collisions with the container walls, *i.e.* $P(\mu)$ is the force per unit surface area exerted over the walls, see §1.5.

Therefore, given a family $\mathcal{E}$ of stationary distributions on the space of microscopic states, we shall call its elements a *statistical ensemble* or simply *ensemble*, after Gibbs, or *monode*, after Boltzmann, see §1.9. If $\mathcal{E}$ is such a family we can associate with every “macroscopic” state $\mu \in \mathcal{E}$ the quantities U, V, K, p (energy, volume, average kinetic energy and average pressure) and we can ask whether the statistical ensemble $\mathcal{E}$ defines a “model of thermodynamics” in which the absolute temperature T can be identified with the average kinetic energy per particle up to a proportionality factor that, to simplify various expressions, is written as $2/3k_B$:

2.1.4
$$T = \frac{2}{3k_B} \frac{K(\mu)}{N} \tag{2.1.4}$$

where k_B is a constant to be determined empirically.

The precise meaning of the locution “defines a model of thermodynamics” has been discussed in Chap.I, (see (1.6.5)); it means that by varying μ in $\mathcal{E}$ and following the variations of U, V, T, p the relation:

2.1.5
$$(\,\mathrm{d}U + p\,\mathrm{d}V)/T = \text{exact differential} \tag{2.1.5}$$

holds. Hence it will be possible, by integrating (2.1.5), to define a function $S(\mu)$ on $\mathcal{E}$ so that the quantities U, V, S, T, p satisfy the relations of classical thermodynamics in which S has the interpretation of “*entropy*”:

2.1.6
$$(\,\mathrm{d}U + p\,\mathrm{d}V)/T = \,\mathrm{d}S, \tag{2.1.6}$$

see (1.5.6).

It is possible, in this way, to associate with each macroscopic state $\mu \in \mathcal{E}$ the quantities U, T, S, p, V and define a “model of thermodynamics”: the statistical ensembles $\mathcal{E}$ that enjoy the latter property (2.1.6) were briefly called by Boltzmann “orthodes”, see §1.6 and §1.9, and therefore we shall refer to (2.1.6) by calling it the *orthodicity property* of the ensemble $\mathcal{E}$.

The existence of important classes of orthodic ensembles was demonstrated by Boltzmann who also provided some *a priori* reasons to expect that his examples should not only give mechanical “models of thermodynamics” but *precisely the* thermodynamics of the system, given to us by the experimental

observations: in this attempt he founded *ergodic theory* and the *Boltzmann equation*, see §1.9.

Therefore the "theory of ensembles" poses three questions:

(1) existence and description of orthodic ensembles

(2) equivalence of the thermodynamics that they describe

(3) comparison of the equations of state computed from the ensembles and the corresponding ones obtained experimentally.

In this chapter we shall consider the two basic ensembles studied by Boltzmann and we shall show their orthodicity, following the lines of Boltzmann, [Bo84].

The *canonical ensemble* (see §1.5) consists of the probability distributions μ on the space of the microscopic states Δ which describe particles roaming in a volume V that, for simplicity, we shall suppose cubic and with perfectly reflecting walls. The probability of a cell is, by definition:

2.1.7
$$\mu(\Delta) = \frac{e^{-\beta E(\Delta)}}{Z(\beta, V)} \tag{2.1.7}$$

with $E(\Delta) = E(\underline{p}, \underline{q})$, $(\underline{p}, \underline{q}) \in \Delta$, being the energy of the microscopical configuration Δ, (2.1.1), and:

2.1.8
$$Z(\beta, V) = \sum_{\Delta} e^{-\beta E(\Delta)} \tag{2.1.8}$$

is a normalization factor Z which will be called *canonical partition function*; the elements μ of the canonical ensemble are therefore parameterized by the volume V and the quantity β.

The *microcanonical ensemble* consists of the probability distributions μ parameterized by the parameters U and V defined by:

2.1.9
$$\mu(\Delta) = \begin{cases} 1/\mathcal{N}(U, V) & \text{if } U - DE \le E(\Delta) \le U \\ 0 & \text{otherwise} \end{cases} \tag{2.1.9}$$

where $\mathcal{N}(U, V)$, called the *microcanonical partition function*, is:

2.1.10
$$\mathcal{N}(U, V) = \sum_{E(D) \in [U - DE, U]} 1 = \left\{ \begin{array}{l} \text{number of cells } \Delta \text{ of} \\ \text{energy } E(\Delta) \in [U - DE, U] \end{array} \right. \tag{2.1.10}$$

where DE is a macroscopic energy, albeit very small compared to U.[1]

[1] Or better compared with $U + BN$ if the energy is bounded below by a stability bound $U_{\min} \ge -BN$; our conventions give energy 0 to configurations in which the N particles are infinitely far apart and with zero speed: hence, if $U_{\min}$ is the minimum (potential) energy then the energy above the minimum energy configuration, the "*ground state*" is $U - U^\circ \ge U + BN$ and $-BN$ is a lower bound for U°.

In other words in the microcanonical ensemble one attributes equal probability to all cells with macroscopic energy U and 0 probability to the others, while in the canonical ensemble one attributes relative probability (also called *weight*) $e^{-\beta U}$ to all cells with microscopic energy U which, however, can take all possible values (*i.e.* all values between the minimum of the potential energy and $+\infty$).

Proving "orthodicity" of the above ensembles means

(a) expressing U, K, p in terms of two parameters (β, v), with $v = V/N, \beta > 0$, in the case of the canonical ensemble, or (u, v) with $u = U/N$, $v = V/N$, in the case of the microcanonical ensemble, and

(b) showing that, defining T equal to $\frac{2}{3k_B}$ times the average kinetic energy per particle, then:

2.1.11
$$(\,\mathrm{d}u + p\,\mathrm{d}v)/T = \text{ exact differential} \tag{2.1.11}$$

as (β, v) or (u, v) vary, respectively.

We shall see that while the canonical ensemble is *already* orthodic in finite volume, the microcanonical ensemble is orthodic "only" in the "*thermodynamic limit*" $N \to \infty$, $U \to \infty$, $V \to \infty$ so that $U/N = u$, $V/N = v$ stay constant (or tend to a constant).

This will be the physically interesting limiting situation, if one keeps in mind the size of N, in real physical systems.

§2.2. Canonical and Microcanonical Ensembles: Orthodicity.

There are many other examples of ensembles which are orthodic at least in the thermodynamic limit. However before proceeding to the discussion of other ensembles and of their equivalence (*i.e.* of their identity as models of thermodynamics) it is convenient to describe how one can check orthodicity of the canonical and microcanonical ensembles. This check is a key to the understanding of Boltzmann's ideas and to the understanding of the mathematical mechanisms that make tractable a problem that at first sight might look formidable.

Consider first the canonical ensemble case (2.1.7), (2.1.8).

The partition sum $Z(\beta, V)$ can be computed, if the cell size $h = \delta p\,\delta q$ is small, as:

2.2.1
$$Z(\beta, V) = \int_{R^{3N}} \int_{\underline{q} \in V^N} e^{-\beta K(\underline{p})} e^{-\beta \Phi(\underline{q})} \frac{\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}}{N!h^{3N}} \tag{2.2.1}$$

where the factor $N!$ takes into account that the N particles are strictly identical and, therefore, indistinguishable as a matter of principle, so that by permuting the N particles one obtains microscopic states described by phase space cells that must be regarded as identical.

We can identify a configuration Δ (*i.e.* a phase space cell, as the two notions coincide having adopted a discrete viewpoint (see §1.1), of the system

by giving the "occupation numbers" n_1 of particles in a small cube C_1 of dimension $(\delta p \delta q)^3 = h^3$, n_2 in the cube C_2, *etc* (the cubes should not be confused with the phase space cells: they are 6–dimensional boxes in the phase space of a single particle). Therefore in (2.2.1) we replace the sum corresponding to (2.1.8) with an integral: *in this way a twofold error is committed*:

(i) an approximation error due to the fact that $E(\underline{p}, \underline{q}) = E(\Delta)$ only at the center of the cell Δ: we shall call this error (*i.e.* the act of confusing the average value in a cell with the actual value at the center) an *analytic error*.

(ii) an error due to the fact that a microscopic configuration $\Delta \subset R^{6N}$ described by giving the numbers $n_1, n_2, \ldots$ is counted in the integral (2.2.1) $N!/n_1!n_2!...$ times instead of $N!$ times. We call this *a combinatorial error*.

Both errors are, obviously, infinitesimal as $h \to 0$ (if one means that both dimensions δp and δq tend to 0 as $h \to 0$, as we always assume). They were neglected by Boltzmann in his analysis since he had no reason to think that the cell size would play any role in his classical world, besides that of allowing one to speak of the "number of configurations" of given energy.

We shall neglect them here as well, postponing their analysis until it will become possible to discuss them *a posteriori* on the basis of theoretical consequences of the theory drawn by neglecting them: hence the theory, once developed, will allow us to evaluate under which physical conditions negligibility of the above errors becomes reasonable.

Anticipating the results of the analysis (see also §1.2; details will be provided in §2.6) these errors will become negligible at "high temperature" and, in the example of a perfect gas ($\Phi(\underline{q}) = 0$), at least for:

2.2.2
$$T > T_q = (m k_B h^{-2} \rho^{-2/3})^{-1} \tag{2.2.2}$$

where $\rho = N/V$, $k_B = 1.38 \times 10^{-16}\, erg{}^\circ K^{-1}$.[2]

The relation (2.2.2) is obtained (summarizing part of the discussion in §1.2) by remarking that the representation of the microscopic states by cells can be consistent only if δp and δq are smaller than the average values of the momentum and of the intermolecular distances.[3]

Since by (2.1.4) the absolute temperature is such that $3k_BT/2$ is the average value of the kinetic energy per particle, *i.e.* the average of $\underline{p}_1^2/2m$, it is clear that the average momentum will have order of magnitude $\overline{p} = \sqrt{m k_B T}$, while the average interparticle distance will be $\overline{q} = \sqrt[3]{V/N} =$

---

[2] The inequality (2.2.2) in the case of hydrogen at normal density and pressure: $m = 3.34 \times 10^{-24} g$, $N = 2.7 \times 10^{19}$ particles in $V = 1\, cm^3$, and choosing $h =$ Planck's constant $= 6.62 \times 10^{-27}\, erg/{}^\circ K$, gives $T_q = 1^\circ K$, very different values for T_q are obtained for other gases; see §1.2, where other necessary conditions for the validity of the approximations are also taken into account.

[3] This is a condition less stringent than the one examined in §1.2, (1.2.18): where $T > T_0$ also imposes the compatibility of the description in terms of cells with the classical microscopic dynamics as a cell permutation.

$\rho^{-1/3}$ and therefore the condition $h \equiv \delta p\ \delta q \simeq \overline{p}\,\overline{q} = \sqrt{mk_BT}\rho^{-1/3}$ follows (necessary but not sufficient) yielding (2.2.2). See §2.6 below for a more detailed analysis.

It is important, however, to keep in mind that when (2.2.2) is not valid, hence the cells sizes cannot be neglected, the very consistency of a cell representation of the microscopic states fails, and the whole theory should be reexamined from scratch. It will appear that in such circumstances *quantum mechanics* becomes important and classical statistical mechanics may lose validity in a fundamental sense.

Making the assumption that (2.2.1) is correct without the necessary analytic and combinatorial corrections and performing the orthodicity analysis is equivalent to setting $h = 0$, *i.e.* to admitting the possibility of infinite precision (simultaneous) measurements of position and momenta of (all) particles.

We can evaluate, following Boltzmann, [Bo84], the thermodynamic quantities in the state described by the canonical distribution μ with parameters β, V.

To simplify notation we shall identify the region V occupied by the system with the measure V of its volume (which we always think of as cubic).

We shall use the fact that in our approximations the probability of finding the system in the microscopic state $\mathrm{d}\underline{p}\ \mathrm{d}\underline{q}$ is $e^{-\beta E(\underline{p},\underline{q})}\ \mathrm{d}\underline{p}\ \mathrm{d}\underline{q}/N!h^{3N}Z(\beta,V)$, so that (2.1.3) become:

2.2.3
$$\begin{aligned}
K &= K(\mu) = \int \Big(\sum_{i=1}^N \frac{\underline{p}_i^2}{2m}\Big) e^{-\beta K(\underline{p})-\beta\Phi(\underline{q})} \frac{\mathrm{d}\underline{p}\ \mathrm{d}\underline{q}}{h^{3N}N!Z(\beta,V)}\\
v &= V/N\\
U &= U(\mu) = \frac{-\partial}{\partial\beta}\log Z(\beta,V)\\
p &= P(\mu) = \sum_Q \frac{N}{Z(\beta,V)}\cdot\\
&\cdot\int_{v>0} e^{-\beta(K(\underline{p})+\Phi(\underline{q}))} 2mv^2\,\frac{s}{S}\,\frac{\mathrm{d}\underline{q}_2\ldots\mathrm{d}\underline{q}_N\,\mathrm{d}\underline{p}_1\ldots\mathrm{d}\underline{p}_N}{h^{3N}N!}
\end{aligned}\tag{2.2.3}$$

where the sum is over the small cubes Q adjacent to the boundary of the box V by a side with area s while $S = \sum_Q s$ is the total area of the container surface and $\underline{q}_1$ is the center of Q (note that $S = 6V^{2/3}$), see §1.4, (1.4.4).

It is not difficult to transform the last of (2.2.3) into a more useful form:

2.2.4
$$p = \beta^{-1}\frac{\partial}{\partial V}\log Z(\beta,V)\,, \tag{2.2.4}$$

the calculation is illustrated in detail in §2.6 below where we also collect other more technical deductions.

At this point we only need a simple direct check. Let:

2.2.5
$$F = -\beta^{-1}\log Z(\beta,V), \qquad S = (U-F)/T \longleftrightarrow F = U - TS \tag{2.2.5}$$

and use (2.2.4),(2.2.3),(2.2.4) to obtain:

2.2.6
$$T = \frac{2}{3k_B}\frac{K(\mu)}{N} = \frac{1}{k_B\beta} \qquad \frac{dT}{T} = -\frac{d\beta}{\beta} \tag{2.2.6}$$

because the integral over $\underline{q}$ in (2.2.3) factorizes and the one over $\underline{p}$ is elementary (*i.e.* gaussian):

2.2.7
$$\begin{aligned} \mathrm{d}F &= (\beta^{-2}\log Z(\beta,V) + \beta^{-1}U)\,\mathrm{d}\beta - \beta^{-1}\frac{\partial}{\partial V}\log Z(\beta,V)\ \mathrm{d}V = \\ &= (F-U)\ \mathrm{d}T/T - p\ \mathrm{d}V = -S\,\mathrm{d}T - p\,\mathrm{d}V \end{aligned} \tag{2.2.7}$$

hence:

2.2.8
$$T\,\mathrm{d}S = \ \mathrm{d}(F+TS) + p\,\mathrm{d}V = \ \mathrm{d}U + p\,\mathrm{d}V \tag{2.2.8}$$

which coincides with (2.1.6).

We also read from the above relations the physical interpretation of the partition function $Z(\beta,V)$: in fact the function $F = -\beta^{-1}\log Z(\beta,V)$ is the *free energy* of thermodynamics.

Equation (2.2.8) shows the orthodicity of the canonical ensemble.

Note that (2.2.8) has been derived *without* any necessity to consider the thermodynamic limit $N\to\infty$, $V\to\infty$, $V/N\to v$, as long as one accepts the approximations leading to (2.2.1) (*i.e.* if the cells size h can be taken as 0 or, more physically, as negligible). This "*unconditional*" validity, for all N and V, *should be regarded as a coincidence*, as the following discussion shows. In the other ensemble cases consideration of the thermodynamic limit is necessary to establish the correct thermodynamic relations between U,T,S,p,ρ,V. In fact, to prove orthodicity of ensembles other than the canonical it is necessary to impose some physically important conditions on the interaction potential energy $\Phi(\underline{q})$: the "stability and temperedness conditions", see below.

In particular the situation is somewhat more involved in the microcanonical ensemble case because in this case it becomes really necessary to consider the thermodynamic limit.

The microcanonical *partition function* is defined in (2.1.9) and, up to the errors already pointed out in the case of the canonical ensembles, it can be written as:

2.2.9
$$\mathcal{N}(U,V) = \int_{J_E}\frac{\mathrm{d}p\,\mathrm{d}q}{h^{3N}N!} \tag{2.2.9}$$

where J_E is the phase space set in which $(U-DE\le E(\underline{p},\underline{q})\le U)$.

The thermodynamic quantities are defined by (2.1.3), and the pressure can be written just as in (2.2.3) with $\mathcal{N}(U,V), 1$ replacing $Z(\beta,V), e^{-\beta(K(\underline{p})+\Phi(\underline{q})}$ respectively, and with the integral extended to the domain $U-DE\le K(\underline{p})+\mathcal{F}(\underline{q})\le U$. In this case U is a parameter defining, together with V, the elements of the ensemble. The temperature is defined as $2/3Nk_B$ times the average kinetic energy per particle. See §1.6.

Also in this case orthodicity is derived by a direct check. Let:

2.2.10
$$S = k_B \, \log \mathcal{N}(U,V) \tag{2.2.10}$$

and let T be $(2/3k_B)$ times the average kinetic energy per particle; one finds:

2.2.11
$$\mathrm{d}S = k_B \left(\frac{1}{\mathcal{N}(U,V)} \frac{\partial \mathcal{N}}{\partial U}(U,V)\ \mathrm{d}U + \frac{1}{\mathcal{N}(U,V)} \frac{\partial \mathcal{N}}{\partial V}(U,V)\ \mathrm{d}V \right) \tag{2.2.11}$$

and we ask whether the right-hand side of (2.2.11) can be written as $(\,\mathrm{d}U + p\ \mathrm{d}V)/T$ with p, V, T defined in (2.1.3), *i.e.* by relations analogous to (2.2.3).
The derivatives of $\mathcal{N}$ can be studied as in the case of the canonical ensemble, and one finds that (2.2.11) can be rewritten, see §2.6, as

2.2.12
$$\mathrm{d}S = k_B \, \frac{3N}{2} \left\langle K(\underline{p})^{-1} \right\rangle \left((1 - \frac{2}{3N})\ \mathrm{d}U + \frac{p\ \mathrm{d}V}{\langle K(\underline{p})\rangle^* \langle K(\underline{p})^{-1}\rangle} \right) \tag{2.2.12}$$

where (if J_E is the domain $U - DE \le E(\underline{p},\underline{q}) \le U$) we have set for α real:

2.2.13
$$\langle K(\underline{p})^\alpha \rangle = \frac{\int_{J_E} K(\underline{p})^\alpha \, \mathrm{d}\underline{p}\ \mathrm{d}\underline{q}/h^{3N}N!}{\int_{J_E}\ \mathrm{d}\underline{p}\ \mathrm{d}\underline{q}/h^{3N}N!} \qquad \alpha \text{ real} \tag{2.2.13}$$

2.2.14
$$\langle K(p)^\alpha \rangle^* = \frac{\int_{J_E, \underline{q}_1 \in dV}\ K(\underline{p})^\alpha \, \mathrm{d}\underline{p}\ \mathrm{d}\underline{q}/h^{3N}N!}{\int_{J_E, \underline{q}_1 \in dV}\ \ \mathrm{d}\underline{p}\ \mathrm{d}\underline{q}/h^{3N}N!} \qquad \alpha \text{ real} \tag{2.2.14}$$

dV being an infinitesimal region (with volume also denoted dV) around V obtained by displacing by a distance η, along the external normal to V, the surface elements of V.
In other words $\langle K(\underline{p})^\alpha\rangle$ is the average value of the α-th power of $K(\underline{p})$ with respect to the considered microcanonical distribution, while $\langle K(\underline{p})\rangle^*$ is the average value of $K(\underline{p})$ with respect to a distribution μ^* obtained by imposing the condition that *one* among the N particles is constrained to be in the region dV around the surface of V. If the relations

2.2.15
$$\langle K(\underline{p})^\alpha \rangle\,,\ \langle K(\underline{p})^\alpha \rangle^* = K(\mu)^\alpha (1 + \vartheta_N) \tag{2.2.15}$$

were valid, with $\vartheta_N \xrightarrow[N\to\infty]{} 0$ and with $K(\mu)$ equal to the average kinetic energy in the microcanonical ensemble, then one could deduce that (2.2.12) becomes, after dividing both sides by N and letting $N \to \infty$ with $\frac{V}{N}, \frac{U}{N}$ constants:

2.2.16
$$\mathrm{d}s = (du + p\, dv)/T\,. \tag{2.2.16}$$

In the microcanonical case one sees from (2.2.12), (2.2.16) that the partition sum directly has the physical meaning of entropy: $S = k_B \log \mathcal{N}(U,V)$. Since $\mathcal{N}(U,V)$ is the number of microscopic states with energy U and allowed volume V (see also §1.4) this is the well-known Boltzmann's relation

expressing entropy as proportional to the logarithm of the number of possible microscopic states with given energy and volume, see [Bo77] and §1.9.

To complete the analysis of the microcanonical ensemble orthodicity it remains to check (2.2.15): as already mentioned one *needs* for this purpose suitable assumptions on the potential energy $\Phi(\underline{q})$.

Such assumptions, which will have an important physical meaning, are:

(a) *Stability*: this means that there is a constant B such that for every configuration $(\underline{q}_1, \ldots, \underline{q}_N) = \underline{q}$:

2.2.17
$$\Phi(\underline{q}) = \sum_{i<j} \varphi(\underline{q}_i - \underline{q}_j) \geq -BN \tag{2.2.17}$$

This property not only says that the potential energy is bounded below (as usual in many mechanical systems) but it also says that its minimum cannot be too small as N grows.

(b) *Temperedness*: there are three constants $C > 0$, $x > 0$, $R > 0$ for which:

2.2.18
$$|\varphi(q - q')| \leq C|q - q'|^{-3-x} \qquad \text{for} \qquad |q - q'| > R \tag{2.2.18}$$

This is essentially a condition that says that "far" particles have "small" interaction: by this hypothesis the interaction energy between a particle and a uniformly filled half-space approaches 0 as the distance between the two tends to ∞. In a large system the macroscopic subsystems have "small" interaction energy (*i.e.* much smaller than the product of the volumes occupied by each). This can be considered as a property of the "short range" of the forces.

Relations (2.2.17),(2.2.18) are not satisfied in the special but *very important* case of systems of charged particles interacting via the Coulomb force: qualitatively the problem really comes only from condition (b) because (a) is satisfied as one thinks that in realistic cases particles have hard cores (however, in spite of this, we shall see that even (a) poses a problem of a quantitative nature as the "obvious" hard cores are often of nuclear size which turns out to be too small for compatibility with the observations). statistical mechanics of systems interacting via Coulomb forces is therefore more delicate than that of systems interacting via phenomenological pair forces with short range (like Lennard-Jones potentials) which mean effective hard cores of atomic size (rather than nuclear size).

Even more delicate is the statistical mechanics of gravitationally interacting particles. We shall see that while systems of charged hard core particles with the property of a *neutral total charge* do obey "normal thermodynamics" the same is *not* true for gravitationally interacting particles (so that we should not expect that a Star obeys the same thermodynamics as a pot of gas, just in case this idea occurred to you).

Equations (2.2.15) are related to the law of large numbers: they say that the variables $\frac{1}{N}K(\underline{p})$, regarded as random variables with a distribution given by an element μ of the microcanonical ensemble or of the corresponding μ^*, see (2.2.15), are variables with a "dispersion that approaches 0" in the limit $N \to \infty$, because the ratio $K(\underline{p})/K(\mu)$ is such that $\langle (K(\underline{p})/K(\mu))^\alpha \rangle$, $\langle (K(\underline{p})/K(\mu))^\alpha \rangle^* \to 1$ for all α; or the fluctuations of $K(\underline{p})^\alpha$, with respect to its average value $\langle K(\underline{p})^\alpha \rangle \simeq K(\mu)^\alpha$, do not have the order of magnitude of $\langle K(\underline{p})^\alpha \rangle$ itself, but are much smaller.

The kinetic energy $K(\underline{p})$ is however a sum of N "almost independent" variables $\frac{p_1^2}{2m}, \ldots, \frac{p_N^2}{2m}$, *i.e.* not really such because they are constrained by $U - DE - \Phi(\underline{q}) \le K(\underline{p}) \le U - \Phi(\underline{q})$). Therefore it is clear that (2.2.15) requires a proof and it does not reduce trivially to the law of large numbers which is formulated for independent variables. We have in fact just discussed which extra assumptions are necessary in order to be able to show the microcanonical ensemble orthodicity.

From a historical viewpoint the above treatment of the canonical ensemble is essentially the same as the original in Boltzmann, [Bo84]; the case of the microcanonical ensemble is somewhat different and more involved: the reason is that in Boltzmann the assumptions (2.2.15) are only implicitly made: in fact Boltzmann studies the problem from a slightly different viewpoint. He considers *a priori* a quantity that he identifies with the amount dQ of heat that the system receives when the microcanonical parameters change by dU, dV. In this way he shows that the microcanonical ensemble is orthodic *even in a finite volume*. This is possible because the definitions of dQ that he uses in the two ensembles are different and in the language used here they are consistent only in the thermodynamic limit (and only if (2.2.15) are assumed). But this is not the moment to attempt a philologically correct treatment of Boltzmann' s ideas (a treatment that is still quite unsatisfactory in the literature, see §1.9).

To conclude this section we can ask how strongly the orthodicity of the canonical and microcanonical ensembles depends upon the hypothesis that (2.2.1) and (2.2.9) are good approximations to the partition sums (as finite sums over cells in phase space), and how strongly the orthodicity depends on the hypothesis that the system consists of only one species of identical particles.

Without exhibiting any analytic calculations we simply say that, in the case that the integrals (2.2.9) or (2.2.1) are replaced by the sums that they are supposed to approximate, orthodicity *must be formulated differently*: in the canonical ensemble one has to interpret β as proportional to the inverse of the absolute temperature while in the case of the microcanonical ensemble one must define the entropy directly via Boltzmann's formula: $S = k_B \log \mathcal{N}(U, V)$.

One obtains in this way two models of thermodynamics, which are models in a sense which is natural although different from the one so far used. Namely in the first case by setting $T = \frac{1}{k_B\beta}$ the expression $(dU + pdV)/T$ is

an exact differential, *but T is no longer proportional to the average kinetic energy*; in the second case setting $T^{-1}=(dU+pdV)/dS$ the quantity T is independent of the transformation that generates the variations dU, dV and the corresponding dS. Furthermore it is possible to prove that the two models of thermodynamics are equivalent, [Ru69].

The important and well-known *universal identification*, [Cl65], [Bo66], between the average kinetic energy with the absolute temperature is *no longer valid*: in view of the role that this identification played in the birth of statistical mechanics and in its developments one should regard this as a shocking major change. See Chap.III for a more detailed analysis of this point.

Therefore the ensembles in which the partition functions are evaluated without the "continuum approximation", valid only when (2.2.2) (or better when (1.2.4), (1.2.5)) hold, can *still* be used for the formal construction of models of thermodynamics.

However, as a consequence of the general considerations following (2.2.2), in such cases it is not clear what the physical meaning of the thermodynamics that is constructed from the mechanical model could be: a physically correct investigation would in fact require, in such situations, using quantum mechanics as a basis for the treatment.

For what concerns the assumption of existence of only one species of particles in the systems considered so far we simply mention that orthodicity does not depend on this assumption. But there are some obvious changes that one has to introduce in the formulation and in the combinatorial factors to be used. As an example we just write the partition function for a general system with N_1 particles of species 1 and mass m_1, N_2 species 2 particles with mass m_2, *etc.* Under the assumption that the cell size can be neglected we have:

2.2.19
$$Z(\beta,V)=\frac{1}{N_1!N_2!\ldots}\int\frac{\mathrm{d}\underline{p}_1\,\mathrm{d}\underline{q}_1}{h^{3N_1}}\frac{\mathrm{d}\underline{p}_2\,\mathrm{d}\underline{q}_2}{h^{3N_2}}\ldots e^{-\beta\sum_\alpha K(\underline{p}_\alpha)-\beta\Phi(\underline{q}_1\ldots)} \tag{2.2.19}$$

and the probability of a microscopic state will be:

2.2.20
$$\prod_\alpha\left(\frac{\mathrm{d}\underline{p}_\alpha\,\mathrm{d}\underline{q}_\alpha}{N_\alpha!h^{3N_\alpha}}\right)e^{-\beta\sum_\alpha K(\underline{p}_\alpha)-\beta\Phi(\underline{q}_1,\ldots)}Z(\beta,V)^{-1}\,. \tag{2.2.20}$$

The natural generalization to this case of the notion of orthodicity is checked in exactly the same way as in the previous case of only one species of particles.

§2.3. Equivalence between Canonical and Microcanonical Ensembles.

In the above study of canonical and microcanonical ensembles Boltzmann's constant appeared several times: it was always denoted by the same symbols, but it was to be regarded as *a priori* different in each case.

In fact this is a *universal constant* $k_B = 1.38 \times 10^{-16}\, erg\, °K^{-1}$.

The logical itinerary leading to the identification of k_B and to showing the equivalence of thermodynamic models described by the orthodic canonical and microcanonical ensembles is discussed in this section.

Suppose first that the molecules do not interact, $\varphi = 0$, *i.e.* consider the microscopic model of a free gas. In this case it is easy to compute explicitly the microcanonical and canonical partition functions, $\mathcal{N}, Z$, in the approximation in which cell size is neglected (see (2.2.1) and (2.2.9)).

One finds, performing the integrals (2.2.1) and (2.2.9) in polar coordinates in momentum space:

2.3.1
$$\begin{aligned} \mathcal{N}(U,V) &= \frac{V^N(\sqrt{2mU}^{3N} - \sqrt{2m(U-DE)}^{3N})}{N!h^{3N}}\frac{\Omega(3N)}{3N} \\ Z(\beta,V) &= \frac{V^N\sqrt{2\pi m\beta^{-1}}^{3N}}{h^{3N}N!} \end{aligned} \tag{2.3.1}$$

where $\Omega(d) = \Gamma(d/2)^{-1}\sqrt{\pi}^d$ is the surface of the d dimensional unit sphere and $\Gamma(x)$ is Euler's gamma function (*i.e.* $\Gamma(x) = (x-1)!$).

The limits of (2.3.1) as $N \to \infty$, $V \to \infty$, with $V/N = v$, $U/N = u$ fixed, are easily studied via Stirling's formula $\Gamma(x+1) = x^x e^{-x}\sqrt{2\pi x}(1 + O(1/x))$, or $N! = N^N e^{-N}\sqrt{2\pi N}(1 + O(1/N))$ and one finds, see §2.1 and (2.2.5),(2.2.10):

2.3.2
$$\begin{aligned} S &= k_B \log \mathcal{N}(U,V) = \\ &= Nk_B(\log\frac{V}{N} + \frac{3}{2}\log\frac{U}{N} + \mathrm{const} + O(\frac{\log N}{N})) \\ F &= -\beta^{-1}\log Z(\beta,V) = \\ &= -N\beta^{-1}(\log\frac{V}{N} - \frac{3}{2}\log\beta + \mathrm{const} + O(\frac{\log N}{N}))\,. \end{aligned} \tag{2.3.2}$$

On the basis of the discussion in §2.2, S has the interpretation of entropy in the microcanonical ensemble and F of free energy, $F = U - TS$, see (2.2.5). Hence we can compute the pressure in both cases:

2.3.3
$$\begin{aligned} \frac{p}{T} &= (\frac{\partial S}{\partial V})_U = k_B\frac{N}{V}\Big(1 + O(\frac{1}{N})\Big) \qquad &\text{microcanonical} \\ p &= -(\frac{\partial F}{\partial V})_\beta = \beta^{-1}\frac{N}{V} = k_B T\, v^{-1} \qquad &\text{canonical} \end{aligned} \tag{2.3.3}$$

If N_A is Avogadro's number ($N_A = 6.0 \times 10^{23}$ molecules per mole) and $N = nN_A$ (with $n =$ number of moles), one sees that (2.3.3) establish that the perfect gas equation of state is $pV = nRT$ in both cases, *provided* the value of k_B is chosen the same in the two cases and provided it has the numerical value:

2.3.4
$$\begin{aligned} k_B = R/N_A = \text{ gas constant}/N_A &= (8.30\,10^7/N_A)\ \mathrm{erg}\,°K^{-1} = \\ &= 1.38 \times 10^{-16}\mathrm{erg}\,°K^{-1}\,. \end{aligned} \tag{2.3.4}$$

The specific heat at constant volume turns out to be, in the thermodynamic limit (after an easy calculation) $\frac{3nR}{2}$: for instance in the canonical ensemble the average total energy, equal to the average total kinetic energy because $\Phi = 0$, is $\frac{3}{2}Nk_BT$ and is volume independent, at fixed N; see (2.2.6) and §2.6 below.

As we see the two thermodynamics defined for the perfect gas by the two microscopic models, canonical and microcanonical, coincide in the thermodynamic limit and they coincide with the experimentally known thermodynamics of a free gas, provided the constant k_B is chosen in both cases as in (2.3.4).

We now ask whether the coincidence of the thermodynamics defined by the two statistical ensembles remains the same also for more general systems.

This is the *problem of equivalence* of the microcanonical and canonical ensembles. It is a fundamental problem because it would be a serious setback for the whole theory if there were different orthodic ensembles predicting different thermodynamics for the same system, *i.e.* different relations among u, v, T, p, s, all compatible with the general laws of classical thermodynamics although different from each other.

We shall see that "in general" for each *given* system there is equivalence (in the thermodynamic limit) between canonical and microcanonical ensembles *if* the constant k_B appearing in the theory of the two ensembles is taken to be the same.

Once equivalence of the thermodynamics, defined either by the canonical or by the microcanonical ensembles corresponding to a given system, has been established we shall ask the further question of whether the constant k_B that appears as proportionality factor between temperature and average kinetic energy per degree of freedom is the same for all other systems, *i.e.* whether the numerical value (2.3.4) is system independent.

The scheme of the proof of equivalence between canonical and microcanonical ensembles, already used by Boltzmann and Gibbs, is the following. Set

2.3.5
$$\mathcal{N}_0(U,V) = \int_{E(\underline{p},\underline{q})\le U} \frac{d\underline{p}d\underline{q}}{h^{3N}N!}\,. \tag{2.3.5}$$

Note that $\mathcal{N}(U,V) = \mathcal{N}_0(U,V) - \mathcal{N}_0(U-DE,V)$ and that the relation between $\mathcal{N}_0$ and Z is simply given by:

2.3.6
$$Z(\beta,V) = \beta\int_{U_{\min}}^{+\infty} \mathrm{d}E e^{-\beta E}\mathcal{N}_0(E,V) \tag{2.3.6}$$

if $U_{\min}$ is the minimum of the energy and if $Z, \mathcal{N}$ are given by (2.2.1), (2.2.9); this is satisfied by integrating (2.3.6) by parts over E; we treat here only the case in which the continuum approximation is accepted, ($h \cong 0$).[4]

[4] But one can check that the ensemble equivalence remains formally valid even if the cell sizes are not neglected provided the orthodicity notion is adapted to the new case as discussed in §2.2.

Hence, see §2.2:

2.3.7
$$F(\beta,V)=-\beta^{-1}\log Z(\beta,V)= \\ =-\beta^{-1}\log\beta-\beta^{-1}\log\int_{U_{\min}}^{\infty}e^{-\beta E}\mathcal{N}_0(E,V)\,\mathrm{d}E\,. \tag{2.3.7}$$

The specific (*i.e.* per particle) thermodynamic quantities in the canonical distribution μ with parameters β, V are, in the thermodynamic limit ($V\to\infty, V/N=v$ fixed):

2.3.8
$$\begin{aligned}
f_c(\beta,v)&=\lim_{N\to\infty}\frac{1}{N}F(\beta,V) && \text{canonical free energy}\\
u_c(\beta,v)&=\lim_{N\to\infty}\frac{U(\mu)}{N}=\frac{\partial\beta f_c}{\partial\beta}(\beta,v) && \text{canonical internal energy}\\
T_c&=\frac{1}{k_B\beta}=\frac{2}{3k_B}\frac{K(\mu)}{N} && \text{canonical absolute temperature}\\
v&=\frac{V}{N} && \text{canonical specific volume}\\
p_c&=\lim_{N\to\infty}P(\mu)=-\frac{\partial f_c}{\partial v}(\beta,v) && \text{canonical pressure}\\
s_c&=\frac{u_c-f_c}{T_c} && \text{canonical entropy}
\end{aligned} \tag{2.3.8}$$

where in expressing u_c, p_c as derivatives of the free energy f_c via (2.2.3), (2.2.4) the operations of differentiation and of limit have been interchanged without discussion, because we proceed heuristically with the aim of exhibiting the essence of the mechanism of equivalence.

The same thermodynamic quantities can be evaluated also in the microcanonical ensemble with parameters U, V; and of course they have an *a priori* different definition:

2.3.9
$$\begin{aligned}
f_m(u_m,v_m)&=-T_m s_m+u_m && \text{microcan. free energy}\\
u_m&=\frac{U(\mu)}{N}=\frac{U}{N} && \text{m.c. internal energy}\\
T_m&=\frac{2}{3k_B}\frac{K(\mu)}{N}=\Big(\frac{\partial s_m}{\partial u_m}(u_m,v_m)\Big)^{-1} && \text{m.c. abs. temperature}\\
v_m&=\frac{V}{N} && \text{m.c. specific volume}\\
p_m&=P(\mu)=T_m\frac{\partial s}{\partial v_m}(u_m,v_m) && \text{m.c. pressure}\\
s_m&=\lim_{N\to\infty}\frac{k_B}{N}\log\big(\mathcal{N}_0(U,V)-\mathcal{N}_0(U-DE,V)\big)= \\
&=\lim_{N\to\infty}\frac{k_B}{N}\log\mathcal{N}_0(U,V) && \text{m.c. entropy}
\end{aligned} \tag{2.3.9}$$

where the expressions for T_m, p_m follow from (2.2.16), the expression for the free energy is the classical thermodynamic definition, while that of the microcanonical entropy requires a digression.

In the theory of the microcanonical ensemble the value of DE is not specified (and it is only subject to the condition that $DE \ll U$ and that DE is a macroscopic quantity, *i.e.* $DE/N \xrightarrow[N\to\infty]{} De > 0$. Nevertheless the theory of the microcanonical ensemble would proceed in the same way even if $DE = U - U_{\min}$, *i.e.* as large as possible, and one would still obtain an orthodic ensemble, hence a model of thermodynamics in which the entropy would have the "new" value $S = k_B \log \mathcal{N}_0(U,V)$.

The function $\overline{s}_m(u,v) = \lim_{N\to\infty} \frac{k_B}{N} \log \mathcal{N}_0(U,V)$ is monotonic non decreasing in u because such is, manifestly, $\mathcal{N}(U,V)$ and in reality one can show that, in the cases we consider (*i.e.* stable and tempered potentials, see (2.2.17),(2.2.18)), it is strictly increasing (see Chap.IV) as we should wish because, if $\overline{s}_m \equiv s_m$, the derivative $(\partial\, \overline{s}_m/\partial u_m)^{-1}$ should be equal to the absolute temperature, which should be positive.

Hence, at the dominant order in $N \to \infty$ and ignoring problems of exchange of limits:

2.3.10
$$\begin{aligned}\mathcal{N}_0(U,V) &= e^{\frac{N}{k_B}\,\overline{s}_m(u,v)}\\ \frac{\mathcal{N}_0(U-DE,V)}{\mathcal{N}_0(U,V)} &= e^{\frac{N}{k_B}(\,\overline{s}_m(u-De,V)-\,\overline{s}_m(u,v))} = e^{-\alpha N}\end{aligned} \tag{2.3.10}$$

and $\alpha > 0$ as a consequence of the strict monotonicity of $\overline{s}_m$ in u, so that the two limits in the last of (2.3.9) coincide and $s_m \equiv \overline{s}_m$. This shows also the equivalence of the various versions of the microcanonical ensemble determined by various choices of $DE = NDe$ with $De > 0$.

Coming back to the equivalence between microcanonical and canonical ensembles we fix the constant k_B in (2.3.8),(2.3.9) to be the *same quantity* and we see that the problem can be formulated as follows: if we establish a correspondence between the canonical state with parameters $\beta = 1/k_BT_c$, $v = v_c$ and the microcanonical state with parameters $u = u_m, v = v_m$ such that $T_c = (k_B\beta)^{-1} = T_m$ and $v_c = v_m$ then all the other quantities with the same "name" (*i.e.* differing only by the label m or c) must coincide. In this way, because of orthodicity, all other thermodynamic quantities must coincide. Hence, if this coincidence really takes place, the two models of thermodynamics defined by the two ensembles will coincide.

The reason why the coincidence takes place is quite simple, if one neglects matters of mathematical rigor and proceeds heuristically. For large N one finds, by (2.3.6) and the first of (2.3.10):

2.3.11
$$\begin{aligned}Z(\beta,v_m) &= \beta\int_{U_{\min}}^{\infty} e^{-\beta E}\mathcal{N}_0(E,V)\,\mathrm{d}E =\\ &= N\beta\int_{U_{\min}}^{\infty} e^{-\beta N u}e^{N s_m(u,v_m)/k_B}\,\mathrm{d}u =\\ &\cong \mathit{const}\,N^{\frac12}\exp\big[N\max_u(-\beta u + \frac{1}{k_B}s_m(u,v_m))\big]\end{aligned} \tag{2.3.11}$$

so that if the maximum is attained at a *unique point* u_0, it must be that u_0 is such that $\beta = \frac{1}{k_B}\frac{\partial s_m}{\partial u}(u_0,v_m)$, because the derivative with respect to u

must vanish in the maximum point u_0. Furthermore:

2.3.12

$$\begin{aligned} u_c = U(\mu)/N = \frac{\int_{U_{\min}}^{\infty} e^{-\beta E}(E/N)\mathcal{N}_0(E,V)\,\mathrm{d}E}{\int_{U_{\min}}^{\infty} e^{-\beta E}\mathcal{N}_0(E,V)\,\mathrm{d}E} = \\ = \frac{\int_{U_{\min}}^{\infty} u\; e^{(-\beta u + s_m(u,v_m)/k_B)N}\,\mathrm{d}u}{\int_{U_{\min}}^{\infty} e^{(-\beta u + s_m(u,v_m)/k_B)N}\,\mathrm{d}u} \to u_0 \end{aligned} \tag{2.3.12}$$

because only the values of $u \cong u_0$ will give a leading contribution to the integrals as $N \to \infty$. Equation (2.3.12) also confirms the physical meaning of u_0: it is the average energy per particle, *i.e.* the internal energy per particle.

Recalling the relation remarked after (2.3.11) between u_0 and β and the fact that $u_c = u_0$, we have:

2.3.13

$$\beta = \frac{1}{k_B}\frac{\partial s_m}{\partial u}(u_c, v_m) = \frac{1}{k_B}\,T_m(u_c, v_m)^{-1} \tag{2.3.13}$$

and choosing $v_c = v_m$ and u_c so that $T_c = T_m(u_m, v_m)$ it follows that $u_c = u_m = u_0$, from the third of (2.3.8), (2.3.9).

It remains to check that $f_m(u_m, v_m) = f_c(\beta, v_c)$; this follows from (2.3.11) which tells us that, for $N \to \infty$:

2.3.14

$$\begin{aligned} f_c(\beta, v_m) &= -\beta^{-1} \max_u(-\beta u + s_m(u, v_m)/k_B) = \\ &= -\beta^{-1}(-\beta u_c + s_m(u_c, v_m)/k_B) = \\ &= (u_c - T_c s_m(u_c, v_m)) = \\ &= (u_m - T_m s_m(u_m, v_m)) = f_m(u_m, v_m) \end{aligned} \tag{2.3.14}$$

because $T_c = T_m$, $u_c = u_m$.

The identity between the free energy, internal energy and absolute temperatures implies (since the ensembles are orthodic, and therefore the usual thermodynamic relations hold) that of the entropies; so that the two ensembles describe the same thermodynamics.

§2.4. Non Equivalence of the Canonical and Microcanonical Ensembles. Phase Transitions. Boltzmann's Constant

The derivation in §2.3 is classical but nonrigorous: it can be made rigorous via a more detailed analysis of the qualitative properties of the functions $s_m(u, v)$ and $f_c(\beta, v)$: the central point of a rigorous proof of equivalence is in showing that $s_m(u, v)$ is "well approximated" (for N large) by $S(U, V)/N$ and, furthermore, it is a convex function of u and a convex function of v, while $f_c(\beta, v)$ is concave in both variables β, v. This implies that the maximum in (2.3.11) is actually reached at a point u_0 or, possibly, in an interval (u_-, u_+) where the function $\beta u - s_m(u, v)$ is constant in u.

A more detailed analysis of the question is postponed to Chap.IV: it is however useful to mention that such an analysis *requires* making use of the stability and temperedness properties of the inter-particles interaction potential φ.

As one can predict from the discussion that we have led, the proof (rigorous or not) of equivalence between the canonical and microcanonical ensembles *no longer works, in general, if the maximum in (2.3.11) is reached on an interval* (u_-, u_+), $u_- < u_+$ rather than at a single point.

By the general properties of concave functions, one can see that this possibility can be realized *only* for exceptional values of β (and precisely for a set of values forming "at most" a denumerable set). This means that, for exceptional values of β, *i.e.* of the temperature, corresponding elements of the canonical and microcanonical ensembles *may be not equivalent.*

Such values of β are exceptional, if they exist at all; therefore it must happen that as close as we wish to one of them, call it $\overline{\beta}$, there exist values β' and β'' which are not exceptional ($\beta'' < \bar{\beta} < \beta'$).

For $\beta = \beta'$ or $\beta = \beta''$ there is equivalence of the corresponding elements of the canonical and microcanonical ensembles; and in one case the internal energy will be $u' < u_-$ and in the other it will be $u'' > u_+$, having denoted by (u_-, u_+) the interval on which the function $-g(u, v) = (-\beta u + s_m(u, v))$ takes its maximum in u for $\beta = \overline{\beta}$, as illustrated in Fig. 2.4.1:[5]

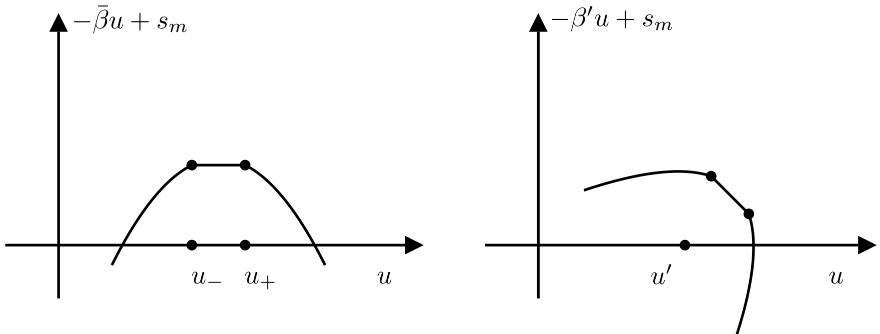


Fig. 2.4.1: Graph of the $-\beta u + s(u, v)$ for different values of β

Hence we see that if for $\beta = \overline{\beta}$ the canonical and microcanonical states are not, or may not be, equivalent then it *must be* that the internal energy $u_c(\beta, v)$ shows a discontinuity jumping from u_- to u_+ when β is varied across $\overline{\beta}$. Consequently also the specific entropy $s_c(\beta, v)$ must show a discontinuity because $f_c(\beta, v) = u_c - T_c s_c$ is necessarily continuous being convex, as mentioned above.

---

[5] Here the graphs will have a continuous first u-derivative if the inverse temperature $\left(\frac{\partial s_m}{\partial u}\right)_v = T_m^{-1}$ is continuous at constant v: this property is usually true but it is nontrivial to prove it. We do not discuss this matter here, but in §4.3 we shall discuss the similar question of the continuity of the pressure as a function of the density at constant temperature. In the Fig. 2.1 we imagine that T_m is continuous (*i.e.* the plateau and the curved parts merge smoothly, "inside the black disks", to first order).

What has just been said, *rather than being an obstacle* to the microscopic formulation of thermodynamics, *shows* the possibility that statistical mechanics can be the *natural frame* in which to study the *phase transition* phenomenon. In fact we see that some of the thermodynamic quantities can have discontinuities in terms of others, exactly of the type empirically observed in phase transition phenomena, where entropy and energy of two coexisting phases are different while the free energy is the same.

Hence cases in which there is no equivalence between corresponding elements of the two ensembles, or more generally when there are corresponding but nonequivalent elements in two orthodic ensembles, can be taken as signaling a phase transition: this is in fact the definition of phase transition that is commonly accepted today.

From the point of view of Physics what happens in a case of nonequivalence between two elements of two orthodic ensembles with corresponding thermodynamic parameters can be clarified by the following considerations. In general the states of an ensemble describe thermodynamic equilibria *but may fail to describe all of them*, *i.e.* all the equilibrium phases (corresponding for instance to a given free energy and temperature, or to given temperature and pressure).

In other words, a given ensemble may be not rich enough to contain among its elements μ the statistical distributions that characterize all the pure phases or their mixtures: usually given a statistical ensemble $\mathcal{E}$ one will find among the $\mu \in \mathcal{E}$ a distribution describing a particular mixture of coexisting phases (if there are more phases possible with the same free energy and temperature) *but it may not contain the distributions describing the other possible phases or mixtrures*.

This is precisely what can be seen to happen in the cases of the canonical and microcanonical ensembles, at least in the few systems in which the theory can be developed until such details are thoroughly brought to light. See Chap.V.

We can therefore conclude, in the case just examined of the canonical and microcanonical ensembles, that they provide equivalent descriptions of the system thermodynamics in the correspondence of the parameter values to which no phase transition is associated. In the other cases the possible nonequivalence *cannot be considered a defect of the theory*, but it can be ascribed to the fact that, when equivalence fails, the elements of the two statistical ensembles that should be equivalent are not because they describe two different phases that may coexist (or different mixtures of coexisting phases).

One of the most interesting problems of statistical mechanics emerges in this way: it is the problem of finding and studying cases of nonequivalence between corresponding elements of the canonical and microcanonical ensembles (or more generally of two orthodic ensembles).

We conclude this section by coming back to the question of the system independence of the Boltzmann constant k_B. The above discussion only shows that the constant k_B appearing in the theory of the canonical ensem-

ble must be the same as that appearing in the theory of the microcanonical ensemble, if one wants the two ensembles to describe the same thermodynamics (apart from the possible existence of phase transitions).

It is, however, easy to give a general argument showing that k_B must be *system independent* and, hence, it has the value given by (2.3.4) computed for the special case of a free gas. The idea is simply that we want our models of thermodynamics to also describe the same thermodynamics for a system that is part of a larger system.

In fact putting into *weak contact*, mechanical and thermal, two systems that are in thermal equilibrium (*i.e.* that have the same temperature) one builds a composite system which, in the canonical ensemble, will be described by a distribution μ with parameters β, v for the first set and by the parameters (β', v') for the second.

We suppose for simplicity that each of the two systems contains only one species of particle. The composite system will then be described by the product distribution $\mu \times \mu'$ because the two systems are independent and their mechanical interaction is supposed negligible (this is the meaning of the phrase "weak mechanical contact").

On the other hand the distribution $\mu \times \mu'$ must be equivalent to a suitable distribution $\overline{\mu}$ for the composite system; a distribution of equilibrium and canonical. In fact we accept that the thermodynamic states of a system can be represented by the elements of an orthodic ensemble.[6] Hence if Δ and Δ' are two cells representing microscopic states of the two systems $\overline{\mu}(\Delta \times \Delta')$ is proportional to $\exp - \overline{\beta}(E(\Delta) + E(\Delta'))$, because the energy of the microscopic state $(\Delta \times \Delta')$ is $E(\Delta) + E(\Delta')$, by the weak mechanical interaction hypothesis. Hence:

2.4.1
$$\exp -\beta E(\Delta) - \beta' E(\Delta')) = \exp - \overline{\beta}(E(\Delta) + E(\Delta')) \tag{2.4.1}$$

for every pair of cells Δ and Δ', hence $\beta = \beta' = \overline{\beta}$.

But $\beta = 1/kT$, $\beta' = 1/k'T$, $\overline{\beta} = 1/\overline{k}T$ where T is the value, common by the assumption of thermal equilibrium, of the temperature in the three systems and $k, k', \overline{k}$ are the three respective values of the constant k_B.

Hence $k = k' = \overline{k}$: *i.e.* k is a universal constant whose actual value k_B can be deduced, as was done above in (2.3.4), from the theory of a single special system, namely that of the free gas which is the easiest to understand.

§2.5. The Grand Canonical Ensemble and Other Orthodic Ensembles

It is easy to see that there exist a *large number of orthodic ensembles.*

•For instance the following generalization of the microcanonical ensemble, with $DE = U - U_{\min}$, *i.e. equal* rather than *small* compared to $U - U_{\min}$

[6] Unless, perhaps, there are phase transitions, an exceptional case that here we shall suppose not to happen as we may imagine changing by a very small amount the thermodynamic parameters of the systems, still keeping thermal equilibrium.

(as assumed in §2.2):

2.5.1
$$\begin{aligned} &\mu(\Delta) = 1/\mathcal{N}_0(U,V) && \text{if}\, E(\Delta) \le U \\ &\mu(\Delta) = 0 && \text{otherwise} \end{aligned} \tag{2.5.1}$$

already considered in §2.3 (after (2.3.9)) is an orthodic ensemble, for the reasons discussed in §2.3.

This ensemble is also called "microcanonical" (although perhaps improperly because this name was introduced for the case $DE = NDe$, $De > 0$, $De \ll U/N$). But this is a somewhat trivial example of a new orthodic ensemble.

•A *different* and wide class of orthodic ensembles can be built by imagining to fix other particles in positions $\underline{q}_1^*, \underline{q}_2^*, \ldots$, and modify $\Phi(\underline{q})$, see (2.1.1), into $\Phi^*(\underline{q})$:

2.5.2
$$\Phi^*(\underline{q}) = \Phi(\underline{q}) + \sum_{i=1}^{N} \sum_j \varphi(\underline{q}_1 - \underline{q}_j^*) \tag{2.5.2}$$

where the sum over $\underline{q}_j^*$ runs over points $\underline{q}_j^*$ *external* to the volume V inside which the system particles are free to roam. The energy Φ^* has the meaning of potential energy of the system in the presence of particles *fixed* at points located *outside* the container.

As the shape or size of the container changes, when we vary V, we imagine to remove the fixed particles whose positions fall into V.

Starting with the potential energy (2.5.2) we form the statistical microcanonical or canonical ensembles with energy function $E(\underline{p}, \underline{q}) = T(\underline{p}) + \Phi^*(\underline{q})$.

If the fixed external particles are distributed reasonably, *e.g.* so that each unit cube only contains a bounded number of fixed particles, or a number slowly increasing with the distance of the cube from the center of V (*i.e.* if the fixed particles are roughly distributed with uniform density) then it can be shown (see Chap.IV) that the ensembles so obtained are orthodic, at least in the thermodynamic limit ($V \to \infty, V/N = v,\ U/N = u$ fixed or $V \to \infty$, $V/N = v$, β fixed (respectively)), provided the interaction potential φ satisfies the stability and temperedness of §2.2. If we do not wish to neglect the cells size then we should apply to such ensembles the comments at the end of §2.2 on the notion of orthodicity.

The above new ensembles are called *microcanonical* or respectively *canonical ensemble "with fixed particle boundary conditions"*. It can be shown that they are equivalent, in the absence of phase transitions, to the usual canonical ensemble, in a sense analogous to that discussed in the previous sections when comparing the canonical and microcanonical ensembles (*i.e.* they generate the same thermodynamics, in the thermodynamic limit). This can be done exactly along the same lines of argument that led to the equivalence between canonical and microcanonical ensembles in §2.3.

Other orthodic ensembles can be obtained by letting N or V vary, *i.e.* by considering simultaneously microscopic states describing systems with different particles numbers N or occupying different volumes V.

•An example which is very important in many applications is the grand canonical ensemble: its elements depend on two parameters $\beta > 0$ and λ. They are probability distributions, on the cells Δ representing the states of an N particle system in a given volume V and with $N = 0, 1, 2, ...$; if $E(\Delta) = E(\underline{p}, \underline{q}) = K(\underline{p}) + \Phi(\underline{q})$ and if $N(\Delta)$ = number of particles in the microscopic state Δ then:

2.5.3
$$\mu(\Delta) = \frac{e^{-\beta\lambda N(\Delta) - \beta E(\Delta)}}{\Xi(\lambda, \beta)} \tag{2.5.3}$$

where the denominator is called the *grand canonical partition function*

2.5.4
$$\Xi(\lambda, \beta) = \sum_{\Delta} e^{-\beta\lambda N(\Delta) - \beta E(\Delta)} \tag{2.5.4}$$

and the thermodynamic limit consists simply in letting $V \to \infty$ keeping λ, β fixed.

•More generally one can replace $\Phi(\underline{q})$ with the potential energy $\Phi^*(\underline{q})$ described in (2.5.2); in this last case we talk about a *grand canonical ensemble "with fixed particle boundary conditions"*.

•A further class of orthodic ensembles is provided by the *pressure ensemble*: it also admits variants with fixed particles boundary conditions. In this ensemble one fixes N but the container V is thought of as variable and susceptible of taking various volume values $V_1 = V,\ V_2 = 2V,\ V_3 = 3V, ...$ *etc*, the shape always remaining cubic.

If Δ is a cell describing a microscopic state with N particles enclosed in a container $V(\Delta)$ and having energy $E(\Delta)$ one defines for each value of the two parameters $p > 0, \beta > 0$:

2.5.5
$$\mu(\Delta) = \frac{e^{-\beta p V(\Delta) - \beta E(\Delta)}}{J(\beta, p)} \tag{2.5.5}$$

where the denominator is called the partition function of the pressure ensemble, and

2.5.6
$$J_N(\beta, p) = \sum_{j=0}^{\infty} \sum_{\Delta:\, V(\Delta) = V_j} e^{-\beta p V(\Delta) - \beta E(\Delta)} \tag{2.5.6}$$

The thermodynamic limit simply consists in letting N tend to infinity.

Remark: One can also imagine taking the containers to be susceptible of assuming a continuum of values, *e.g.* any volume V keeping it, however, homothetic to a reference shape V^0, *e.g.* to a unit cube); in this case the

sum over the volumes should be replaced by an integral $(V_0)^{-1}\int_0^\infty dV$. A simple application of the pressure ensemble will be found in §5.8.

The theory of the grand canonical and pressure ensembles, as well as the theory of the various ensembles with fixed particle boundary conditions, can be developed by showing their equivalence with the canonical, or microcanonical ensemble, and this can be done via the method of the "maximum value" that we have described in §2.3: it will work if the interparticle potential φ satisfies the stability and temperedness conditions, see (2.2.17) and (2.2.18), [Fi64],[Ru69].

As a further example of the maximum value method, of very common use in statistical mechanics, we deduce some of the properties of the grand canonical ensemble from the corresponding properties of the canonical ensemble and show their equivalence. Again we proceed heuristically, by ignoring problems of mathematical rigor.

If μ is a generic element of the grand canonical ensemble corresponding to the parameters λ,β one has, see (2.3.7) and the first of (2.3.8):

2.5.7

$$\begin{aligned}\Xi(\lambda,\beta) &= \sum_{N=0}^{\infty} e^{-\beta\lambda N} \sum_{\Delta, N(\Delta)=N} e^{-\beta E(\Delta)} = \\ &= \sum_{N=0}^{\infty} e^{-\beta\lambda N} Z_N(\beta,V) \cong \sum_{N=0}^{\infty} e^{-\beta\lambda N} e^{-\beta N f_c(\beta,V/N)} = \\ &= \sum_{N=0}^{\infty} \exp V(-\beta\lambda v^{-1} - \beta v^{-1} f_c(\beta,v))\end{aligned} \tag{2.5.7}$$

where in the last sum $v = V/N$ and $Z_N(\beta,V)$ is the canonical partition function for N particles in the volume V and with temperature $T = 1/k_B\beta$. Hence for $V \to \infty$, and if v_0 is the value where the function $-\beta\lambda v^{-1} - \beta v^{-1} f_c(\beta,v)$, of the variable v, attains its maximum we find:

2.5.8

$$\lim_{V\to\infty} (1/V)\log\Xi(\beta,\lambda) = -\beta\lambda v_0^{-1} - \beta v_0^{-1} f_c(\beta,v_0) \tag{2.5.8}$$

assuming that the maximum point v_0 is unique. Here v_0 satisfies (if one recalls that by (2.3.8) $p_c = -\frac{\partial f_c}{\partial v}(\beta,v)$)

2.5.9

$$\frac{\partial}{\partial v^{-1}}(\beta\lambda v^{-1} + \beta v^{-1} f_c(\beta,v)\big|_{v=v_0} = 0 \to \lambda + f_c(\beta,v_0) + v_0 p_c(\beta,v_0) = 0 \tag{2.5.9}$$

On the other hand v_0 has the interpretation of grand canonical specific volume v_g because:

2.5.10

$$\begin{aligned}v_g^{-1} = N(\mu)/V &= \frac{\sum_{N=0}^\infty \frac{N}{V} e^{-\beta\lambda N} Z_N(\beta,V)}{\sum_{N=0}^\infty e^{-\beta\lambda N} Z_N(\beta,V)} = \\ &= \frac{\sum_{N=0}^\infty N V^{-1} e^{-(\beta\lambda v^{-1} + \beta v^{-1} f_c(\beta,v))V}}{\sum_{N=0}^\infty e^{-(\beta\lambda v^{-1} + \beta v^{-1} f_c(\beta,v))V}} \xrightarrow[V\to\infty]{} \\ &\xrightarrow[V\to\infty]{} v_0^{-1}\end{aligned} \tag{2.5.10}$$

due to the maximum of $-\beta\lambda v^{-1} - \beta v^{-1} f_c(\beta, v)$ being isolated and at the point v_0.

Therefore from (2.5.9), and classical thermodynamics, one finds the physical meaning of λ:

2.5.11
$$-\lambda N = F + pV = U - TS + pV = N\left(f_c(\beta, v_g) + v_g p_c(\beta, v_g)\right) \tag{2.5.11}$$

i.e. $-\lambda N$ is the *Gibbs potential* corresponding to the parameters (β, v_g). Furthermore from(2.5.8) one finds that

2.5.12
$$\lim_{V\to\infty} \frac{1}{V} \log \Xi(\lambda, \beta) = \beta p_c(\beta, v_g) \tag{2.5.12}$$

i.e. the grand canonical partition function is directly related to the canonical pressure associated with the parameters (β, v_g).

This suggests that the grand canonical and the canonical ensembles are equivalent if the elements with parameters (λ, β) and (β, v_g) with $v_g = v_0$, see (2.5.9), are put in correspondence (*i.e.* are thought to describe the same macroscopic state). This can be checked by setting, see (2.1.3):

2.5.13
$$\begin{aligned}
u_g &= \lim_{V\to\infty} \sum_{\Delta} \mu(\Delta) U(\Delta)/N(\Delta)\\
T_g &= \lim_{V\to\infty} (2/3k_B) \sum_{\Delta} \mu(\Delta) K(\Delta)/N(\Delta)\\
v_g &= \lim_{V\to\infty} \sum_{\Delta} \mu(\Delta) V/N(\Delta)\\
p_g &= \lim_{V\to\infty} \sum_{\Delta} \mu(\Delta) p(\Delta)\\
s_g &= (u_g - \lim \beta^{-1}(1/V) \log \Xi(\beta, \lambda))/T_g
\end{aligned} \tag{2.5.13}$$

and by showing the identity between the above quantities computed in the grand canonical ensemble with parameters (λ, β) and the quantities with the same name computed in the canonical ensemble with parameters (β, v_g).

Using the fact that $T_c(\beta, v) \equiv \frac{1}{k_B\beta}$ and

2.5.14
$$T_g = \lim_{V\to\infty} \frac{\sum_{N=0}^{\infty} e^{-\beta\lambda N} Z_N(\beta, V) T_c\left(\beta, \frac{V}{N}\right)}{\sum_{N=0}^{\infty} e^{-\beta\lambda N} Z_N(\beta, V)} \equiv \frac{1}{k_B \beta} \tag{2.5.14}$$

we see that, for the same reason used in deriving $v_g = v_0$ in (2.5.10):

2.5.15
$$\begin{aligned}
u_g &= \lim \frac{\sum_N e^{-\beta\lambda N} Z_N(\beta, V)\, u_c\left(\beta, \frac{V}{N}\right)}{\sum_N e^{-\beta\lambda N} Z_N(\beta, V)} = u_c(\beta, v_0)\\
p_g &= \lim \frac{\sum_N e^{-\beta\lambda N} Z_N(\beta, V)\, p_c\left(\beta, \frac{V}{N}\right)}{\sum_N e^{-\beta\lambda N} Z_N(\beta, V)} = p_c(\beta, v_g)\,,
\end{aligned} \tag{2.5.15}$$

(2.5.14), (2.5.15), (2.5.12) clearly show that all grand canonical thermodynamic quantities coincide with the corresponding canonical quantities.

As in the previous case of canonical versus microcanonical ensembles the above analysis is not rigorous because it involves various interchanges of limits and, furthermore, it presupposes that $(-\lambda v^{-1} - v^{-1} f(\beta, v))$ has a unique isolated maximum (as a function of v_0) but under the assumptions of stability and temperedness of §2.2, (2.2.17), (2.2.18), the problems of mathematical rigor can again be solved, see Chap.IV.

From the theory of the canonical ensemble, which we shall elaborate on in more detail in Chap.IV, it follows also that the function $-v^{-1} f(\beta, v)$ is convex in β and in v^{-1} so that, with "few" exceptional values (*i.e.* at most a denumerable family of values of λ),[7] the function $-(\lambda v^{-1} + v^{-1} f_c(\beta, v))$ has a unique maximum point as a function of v^{-1}. For λ outside this exceptional set there is complete equivalence between thermodynamics of the equilibrium states in terms of the elements of the canonical and grand canonical ensembles.

For the other values of λ (if any) the function $-(\lambda v^{-1} + v^{-1} f(\beta, v))$ takes the maximum value in an interval (v_-, v_+), as implied by the general properties of concave functions, see the Fig. 2.4.1, in such cases the descriptions of states in terms of canonical or grand canonical distributions may be nonequivalent. But the interpretation of the nonequivalence is again that of occurrence of a phase transition: non equivalence has to be interpreted by attributing it to the fact that the distributions in question describe two different equilibria that can coexist in thermodynamic equilibrium (*i.e.* they both have the same temperature and pressure, but different specific volume, entropy, *etc*) in the same sense as discussed in §2.4.

One of the main results of statistical mechanics which we wish to quote with more detail has been that of showing that, at least in many interesting cases, there is complete equivalence between the ensembles which will be called here *enlarged ensembles*: such ensembles are obtained from a given ensemble of stationary distributions (like the canonical, microcanonical or grand canonical) by adding to it all the distributions with boundary conditions of (arbitrarily) fixed external particles.

In such larger ensembles it may still happen that two given states, corresponding to the same values of temperature and pressure, may have different

---

[7] Of course a denumerable set of values has zero length but it might be quite large in other senses: for instance it could be dense! Therefore this easy way of saying that phase transitions are "rare" is very unsatisfactory. But a more detailed analysis is very difficult and perhaps impossible at the level of generality in which we are discussing the matter. More detailed statements, *e.g.* that the discontinuities in λ take place at finitely many values of β and in the plane λ, β they occur on smooth lines, can be derived only when considering very special cases; see Chap.VI. One can ask whether the pressure $-\left(\frac{\partial f_m}{\partial v}\right)_\beta = p$ is a continuous function of v: if so $\lambda v^{-1} + v^{-1} f_m(\beta, v)$ will have continuous v-derivative and we could draw a figure similar to Fig. 2.4.1 in §2.4. Also in this case, see the corresponding comment about the temperature in the microcanonical ensemble in footnote 5 in §2.4, it is nontrivial to show that p is a continuous function of the specific volume at constant temperature: in §4.3 we shall discuss this in more detail, in the case of hard core systems.

averages for other thermodynamic quantities (energy, entropy, specific volume, *etc*) but it will happen that for every element of one ensemble there is another in the other ensemble describing exactly the same macroscopic state and thermodynamics; *i.e.* that associates the same values to all thermodynamic quantities and even the same relative probability distribution to the most probable microscopic states.

In other words we can say that even the phase transition phenomenon can be studied in an enlarged ensemble without worrying that in this way one may "miss" some phases, because the enlarged statistical ensembles are often rich enough to contain all possible phases and their mixtures.

This is the way to understand the nonequivalence between ensembles which normally describe correctly the thermodynamics of the system (*i.e.* that are orthodic). One can, in fact, think that, given a state of a system, one can look at an ideal macroscopic region which however occupies a volume *less* than the total; then the system in this volume can be regarded as a system in equilibrium with its surroundings. Such a system will have a fixed volume but a variable particle number. One would thus describe it naturally in the grand canonical ensemble; however the particles in the system have interactions with the identical particles that are outside the ideal volume selected.

If one imagines taking a picture of the configuration one will see a sample of the configuration in the inner volume and one in the outer volume. Taking the highly imaginative step of collecting only the pictures in which the external configuration is the same we should still see statistically the same state inside the ideal box: this means that the state we see in the ideal box is determined by the state of the particles outside it provided they are chosen in a configuration *"typical"* for the state that is being considered. If this is so we can expect that the grand canonical ensemble with fixed particle boundary conditions can describe *all* possible states.

When there is more than one equilibrium state we can describe them by selecting at random a configuration of the system and forming the grand canonical distribution in a large volume with boundary conditions given by the selected configuration of particles.

Note that from this viewpoint the phenomenon of phase transitions appears as an *instability* of the thermodynamic properties of a system with respect to variations of boundary conditions: for instance keeping the same temperature and pressure but changing boundary conditions one can obtain different values for intensive thermodynamic quantities like the specific energy, the specific entropy, the specific volume, *etc*, *i.e.* by changing the forces that act near the boundary of our system we can change the macroscopic state even if the system is very large (hence the boundary is far and relatively small compared to the volume).

In a sense this is a further manifestation of the richness of statistical mechanics: such a complex phenomenon as a phase transition seems to find its natural theoretical setting, and the bases for its analysis, in the theory of (orthodic) ensembles.

In a given statistical ensemble macroscopic thermodynamic quantities appear either as parameters of the ensemble, *e.g.* u, v in the microcanonical case, β, v in the canonical and λ, β in the grand canonical, or as quantities directly related to the ensemble partition function such as the entropy, the free energy or the pressure in the above three cases, or *they are related to derivatives of the partition function* like the temperature, the pressure and the energy in the three cases, respectively.

It can be shown, see Chap.IV, that the first two types of quantities do not depend on the boundary conditions (imagining the latter to be taken as fixed particle boundary conditions). Hence a way of searching for phase transitions in particular models (*i.e.* in systems obtained by assuming specific choices for the interaction potentials) is to look for parameter values of the chosen statistical ensemble (*e.g.* u, v in the microcanonical case, β, v in the canonical and λ, β in the grand canonical, β, p in the pressure ensemble case) in correspondence of which the thermodynamic function associated with the partition function is *not* differentiable, [Ru69].

This is a method that has become classical: it has, however, the defect of not directly providing a microscopic description of the equilibrium states describing the different possible phases. It determines the location of the phase transition, in the thermodynamic parameter space of the ensemble adopted for the analysis: but it does not analyze the characteristic physical peculiarities of the possible microscopic distributions that describe the various phases.

On the other hand, the study of the boundary condition dependence of the equilibrium states of an "enlarged ensemble" is potentially richer in information and it can lead to a microscopic description of the phase transition and phase coexistence phenomena, because each state of thermodynamic equilibrium is described in detail by a probability distribution of its microscopic configurations. The best understanding is obtained by examining in detail some simple case (there are not, however, many cases in which the above statements can be followed and checked in detail): this will be the theme of Chap.VI, where the Ising model for ferromagnetism will be discussed in connection with the spontaneous magnetization phase transition, and Chap.VII where other simple models are discussed.

Thus we have met two possible definitions of phase transitions. *A system shows a phase transition if a derivative of the thermodynamic function associated with the partition function of an orthodic ensemble has a discontinuity as a function of the parameters describing the elements of the ensemble.* Alternatively: *a phase transition occurs if by changing the boundary conditions in the elements of an enlarged orthodic ensemble one can change, in correspondence of suitable values of the parameters describing the elements of the ensemble, bulk properties of the system.*

We conclude by mentioning that if one develops the thermodynamics model associated with the pressure ensemble along the above lines one easily checks

the equivalence between pressure ensemble and canonical ensemble and one finds that the Gibbs potential $\lambda = pv + u - Ts$ is related to the pressure ensemble partition function by:

2.5.16
$$-\beta\lambda = \lim_{N\to\infty} \frac{1}{N} \log J_N(\beta, p)\,. \tag{2.5.16}$$

In fact one deduces, from the definition (recalling that V^0 is the volume of the reference container box, see the remark following (2.5.6):

2.5.17
$$J(\beta, p) = \int \frac{dV}{V^0} e^{-\beta p V} Z_V(\beta, N) \tag{2.5.17}$$

that $-\frac{1}{N}\log J_N(\beta,p) \xrightarrow[N\to\infty]{} \beta\lambda = \min_v(\beta p\, v + \beta f_c(\beta, v)$ so that the minimum is at v_p such that $\beta p = -\frac{\partial\varphi_c}{\partial v}(\beta, v_p)$. Hence we see that p is identified with the pressure and, by (2.5.6), $\beta\lambda$ with the Gibbs potential.

§2.6. Some Technical Aspects

Some details concerning the derivation of the mathematical identities used in §2.1,§2.2 and some related matters will be provided in this section: namely we shall derive equations (2.2.4), (2.2.12), (2.2.2). The reader wishing to delve into the subject can begin by consulting [Fi64], [Mi68], [Ru69], [La72], [LL72].

(1) It is certainly worth commenting on the step from the last of (2.2.3) to (2.2.4).

In the last of (2.2.3) one can make use of the independence of the integrals performed with respect to the variables $\underline{p}$ from those performed with respect to the variables $\underline{q}$, and the symmetry of the p-components of the integrand.

In this way one can replace $2mNv^2$ by mNv^2 and eliminate the condition $v > 0$; then mNv^2 can be replaced by $N\underline{p}_1^2/3m$, thus taking advantage of the symmetry of the $\underline{p}_1$ dependence in the three components of $\underline{p}_1$.

Hence one can replace the integral on $\underline{p}_1$ that in (2.2.3) is:

2.6.1
$$\int_{v>0} e^{-\beta\underline{p}_1^2/2m} 2mv^2\, \mathrm{d}\underline{p}_1 \quad \text{with} \quad \int e^{-\beta\underline{p}_1^2/2m} \frac{\underline{p}_1^2}{3m}\, \mathrm{d}\underline{p}_1 \tag{2.6.1}$$

and a simple calculation shows that:

2.6.2
$$\int e^{-\beta\underline{p}_1^2/2m} \frac{\underline{p}_1^2}{3m}\, \mathrm{d}\underline{p}_1 = \frac{1}{\beta} \int e^{-\beta\underline{p}_1^2/2m}\, \mathrm{d}\underline{p}_1 \tag{2.6.2}$$

so that:

2.6.3
$$p = \sum_Q \frac{N}{\beta} \frac{s}{S} \frac{1}{Z(\beta, V)} \int_{\underline{q}_2,\ldots,\underline{q}_N \in V^N} e^{-\beta E(\underline{p},\underline{q})} \frac{\mathrm{d}\underline{q}_2 \ldots\, \mathrm{d}\underline{q}_N\, \mathrm{d}\underline{p}_1 \ldots \mathrm{d}\underline{p}_N}{h^{3N} N!} \tag{2.6.3}$$

and the point $\underline{q}_1$ is, in each addend of (2.6.3) localized in Q (which is supposed so small that it has no importance where $\underline{q}_1$ exactly is inside Q).

We now imagine varying V from V to $V + \mathrm{d}V$, increasing the volume by displacing, along the outer normal, by η *every* area element of the surface.

We see that $\log Z(\beta, V)$ varies, since $\mathrm{d}V = S\eta$, by:

2.6.4

$$\begin{aligned}
&\mathrm{d}\log Z(\beta,V) = \\
&= \sum_Q \frac{Ns\eta}{Z(\beta,V)} \int_{\underline{q}_2,\ldots,\underline{q}_N \in V} e^{-\beta E(\underline{p},\underline{q})} \frac{\mathrm{d}\underline{q}_2 \ldots \mathrm{d}\underline{q}_N \; \mathrm{d}\underline{p}_1 \ldots \mathrm{d}\underline{p}_N}{h^{3N} N!} \\
&= \mathrm{d}V \sum_Q \frac{N}{Z(\beta,V)} \frac{s}{S} \int_{\underline{q}_2,\ldots,\underline{q}_N \in V} e^{-\beta E(\underline{p},\underline{q})} \frac{\mathrm{d}\underline{q}_2 \ldots \mathrm{d}\underline{q}_N \; \mathrm{d}\underline{p}_1 \ldots \mathrm{d}\underline{p}_N}{h^{3N} N!}
\end{aligned} \tag{2.6.4}$$

which, comparing with (2.6.3) proves (2.2.4).

(2) Another deduction calling for further details is the step from (2.2.11) to (2.2.12).

By proceeding as in the derivation of (2.6.3) one finds $p\,\mathrm{d}V$ starting from the expression for p as the average $P(\mu)$ with respect to the microcanonical distribution with parameters (U, V). Denoting by $\int^*$ the integral over $(\underline{p}, \underline{q})$ extended to the domain of the $(\underline{p}, \underline{q})$ such that $E - DE \le E(\underline{p}, \underline{q}) \le E$ *and*, at the same time, $\underline{q}_1 \in \mathrm{d}V = \cup_Q Q$:

2.6.5

$$p\ \mathrm{d}V = \frac{N}{\mathcal{N}(U,V)} \int^* \frac{2}{3} \frac{\underline{p}_1^2}{2m} \frac{\mathrm{d}\underline{p}\ \mathrm{d}\underline{q}}{h^{3N} N!} = \frac{2}{3\mathcal{N}(U,V)} \int^* K(\underline{p}) \frac{\mathrm{d}\underline{p}\ \mathrm{d}\underline{q}}{h^{3N} N!} \tag{2.6.5}$$

having again used in the last step the symmetry of $K(\underline{p})$ in $\underline{p}_1, \ldots, \underline{p}_N$ (to eliminate the factor N) and having written $\sum_Q s\eta \cdot = \int_{\underline{q}_1 \in dV} \cdot \, \mathrm{d}\underline{q}_1$ to obtain a more elegant form (formally eliminating the summations over Q naturally appearing, in conformity to its definition, in the expression of the pressure).

To connect (2.6.5) with the derivatives of $\mathcal{N}$ we have to make more explicit the dependence of $\mathcal{N}$ on U, by evaluating exactly the integral (2.2.9) on the $\underline{p}$ variables in polar coordinates (which is an elementary integral).

If $\Omega(3N)$ is the surface of the unit sphere in $3N$ dimensions and if we set $w(U, \underline{q}) = \sqrt{2m(U - \Phi(\underline{q}))}$, we deduce

2.6.6

$$\mathcal{N}(U,V) = \int \frac{\mathrm{d}\underline{q}}{h^{3N} N!} \Big(w(U,\underline{q})^{3N} - w(U-DE,\underline{q})^{3N} \Big) \frac{\Omega(3N)}{3N} \tag{2.6.6}$$

hence:

2.6.7

$$\frac{\partial \mathcal{N}}{\partial U} = \int_{V^N} \frac{\mathrm{d}\underline{q}}{h^{3N} N!} \frac{3N}{2} 2m \Big(w(U,\underline{q})^{3N-2} - w(U-DE,\underline{q})^{3N-2} \Big) \frac{\Omega(3N)}{3N} \tag{2.6.7}$$

and coming back to the original coordinates

2.6.8 $$\frac{1}{\mathcal{N}}\frac{\partial\mathcal{N}}{\partial U}=\frac{3N}{2}\frac{1}{\mathcal{N}}\int\frac{\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}}{h^{3N}N!}\frac{3N-2}{3N}\frac{1}{K(\underline{p})}=\frac{3N}{2}\Big(1-\frac{2}{3N}\Big)\langle K(\underline{p})^{-1}\rangle \quad (2.6.8)$$

where $\langle K(\underline{p})^\alpha\rangle$ is defined by (2.2.13) and the integral (2.6.8) is extended to the domain in which $(U-DE\le E(p,q)\le U)$.

If, instead, one had proceeded as in the derivation of (2.2.4) in the canonical ensemble case (see (2.6.4) above) one would have found:

2.6.9 $$\begin{aligned}\frac{1}{\mathcal{N}}\frac{\partial\mathcal{N}}{\partial V}\,\mathrm{d}V&=\frac{N}{\mathcal{N}}\int^*\frac{\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}}{h^{3N}N!}=\\&=\frac{2N}{3}\frac{\Big(\int^* K(\underline{p})\frac{\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}}{h^{3N}N!}/\mathcal{N}\Big)}{\Big(\frac{2}{3}\int^* K(\underline{p})\frac{\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}}{h^{3N}N!}\Big/\int^*\frac{\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}}{h^{3N}N!}\Big)}\end{aligned}\qquad(2.6.9)$$

where in the last step we multiply and divide by the same quantity and we use the notation $\int^*$ in (2.6.5). Then (2.6.9) and (2.6.5) imply

2.6.10 $$\frac{1}{\mathcal{N}}\frac{\partial\mathcal{N}}{\partial V}\,\mathrm{d}V=N\frac{p\,\mathrm{d}V}{\frac{2}{3}\langle K(\underline{p})\rangle^*}\qquad(2.6.10)$$

where $\langle K(\underline{p})^\alpha\rangle^*$ is defined in (2.2.14).

Relation (2.2.12) now follows from (2.2.11), (2.6.8) and (2.6.10).

(3) We finally deduce (2.2.2) in the simple case (considered in §2.2) of a perfect gas, $\Phi=0$.

If we imagine dividing the six-dimensional phase space describing the individual particles of the system, into cells C having the form:

$$C=\quad\text{set of the }\ (\underline{p},\underline{q})\ \text{ in }\ R^6\quad\text{such that:}$$

2.6.11 $$\begin{cases}k'_\alpha\delta q-\delta q/2\le q_\alpha\le k'_\alpha\delta q+\delta q/2\\ k_\alpha\delta p-\delta p/2\le p_\alpha\le k_\alpha\delta p+\delta p/2\end{cases}\qquad\alpha=1,2,3\qquad(2.6.11)$$

and $\underline{k},\underline{k}'$ are two integer components vectors; it follows that the energy of a single particle located in C is $\varepsilon(C)$:

2.6.12 $$\varepsilon(C)=\sum_{\alpha=1}^{3}\frac{1}{2m}\underline{k}_\alpha^2\delta p^2\,.\qquad(2.6.12)$$

Furthermore a microscopic state Δ of the system can be assigned by giving the *occupation numbers* n_C for each cell: they tell us how many particles occupy a given cell. Then, without combinatorial or analytical errors (see §2.2):

2.6.13 $$Z(\beta,V)=\sum_{\substack{n_C\ge0\\ \sum_C n_C=N}}e^{-\beta\sum_C n_C\varepsilon(C)}\,.\qquad(2.6.13)$$

Taking $L = V^{1/3}$ to be the side of the container and calculating, instead, the expression afflicted by obvious combinatorial errors:

$$Z(\beta,V) = \sum_{\substack{n_C\ge 0\\ \sum_C n_C=N}} \frac{1}{(\prod_C n_C!)} e^{-\beta\sum_C n_C\varepsilon(C)} =$$

$$= \frac{1}{N!}\Big(\sum_C e^{-\beta\varepsilon(C)}\Big)^N = \frac{1}{N!}\Big(\frac{L}{\delta q}\sum_{k=-\infty}^{+\infty} e^{-\beta k^2(\delta p)^2/2m}\Big)^{3N} =$$

2.6.14 $$= \frac{1}{N!}\Big(\frac{L\delta p}{h}\sum_{-\infty}^{+\infty} e^{-\beta k^2(\delta p)^2/2m}\Big)^{3N} = \qquad (2.6.14)$$

$$= \frac{1}{N!}\frac{V^N\sqrt{2m\beta^{-1}}^{3N}}{h^{3N}}\Big(\frac{1}{\sqrt{2m\beta^{-1}}}\delta p\sum_{k=-\infty}^{+\infty} e^{-\beta k^2(\delta p)^2/2m}\Big)^{3N}$$

which then leads to (2.2.1) (*i.e.* to the second of (2.3.1), since $\varphi = 0$) if $h \cong 0$ and if one approximates the sum in the last member of (2.6.14) with the corresponding integral $\sqrt{\frac{\beta}{2m}}\int e^{-\beta p^2/2m}\, dp$ (committing in this way also the above described analytic error).

To compare (2.6.13) with (2.6.14) or (2.2.1) it is necessary to decide whether the values of n_C that give the main contribution to (2.6.13) are those for which $n_C = 0, 1$ (and in this case (2.6.13),(2.6.14) are good approximations of each other as well as of (2.2.1) because the factor $n_C!$ has value 1 in the majority of cases).

We must therefore compute the average value $\overline{n}_C$ of the quantity n_C with respect to the canonical distribution and the consistency condition, *i.e.* the condition of negligibility of the combinatorial error will be $\overline{n}_C \ll 1$.

In the canonical ensemble, by definition (2.1.7), the probability of finding a particle, with known position, with momentum in $d\underline{p}$ is the Maxwell-Boltzmann law:

2.6.15 $$\frac{e^{-\beta\underline{p}^2/2m}\, d\underline{p}}{(\sqrt{2\pi m\beta^{-1}})^3}\,; \qquad (2.6.15)$$

hence if $\rho = N/V$ is the system density we shall find:

2.6.16 $$\overline{n}_C = \rho(\delta q)^3 e^{-\beta\underline{p}^2/2m}\frac{(\delta p)^3}{\sqrt{2\pi m\beta^{-1}}^3} \le \frac{\rho h^3}{\sqrt{2\pi m\beta^{-1}}^3} \qquad (2.6.16)$$

so that $\overline{n}_C \ll 1$, for all cells C, if $T > T_q$ with T_q given by (2.2.2).

It is clear that the error that we call "analytic error" will be negligible if $\sqrt{\beta/2m}\,\delta p \ll 1$. In the present context we did not fix separately δp and δq: nevertheless δq should be certainly chosen so that $\delta q > \rho^{-1/3}$ = average inter-particles distance, otherwise it would not make sense to think of the system as built with particles as separate entities defined in the system. With this choice of δq, from $\delta p\ \delta q = h$, it follows that $\delta p = h/\rho^{-1/3}$ and one sees that the condition $\sqrt{\beta/2m}\delta p \ll 1$ is the same as (2.2.2).

Chapter III:

Equipartition and Critique

.

§3.1. Equipartition and Other Paradoxes and Applications of Statistical Mechanics

One of the most well-known consequences of classical statistical mechanics is the *principle of equipartition of energy*: somewhat less well known is that this principle, after some shining initial successes, reveals itself as the sign of statistical mechanics inadequacy to solve important problems that fall into its domain. Likewise other well-known important consequences of statistical mechanics are affected by serious paradoxes and theoretical problems.
Here we shall illustrate some significant examples.

(I) The Free Gases Specific Heats

By using the canonical ensemble and assuming that the cells size h is very small (see §1.2) one easily computes the internal energy for a general model in which each particle has ℓ degrees of freedom and does not interact with the others.
The ℓ degrees of freedom describe the three baricentric degrees of freedom plus the $\ell - 3$ internal degrees of freedom of the molecule internal motion.
One supposes that energy is a quadratic form in the ℓ conjugated momenta $p_1, p_2, \ldots, p_\ell$ and, possibly, in some of the internal position coordinates:

3.1.1 $$E(p,q) = \sum_{j=1}^{3} \frac{p_j^2}{2m} + \sum_{j=4}^{\ell_0} \frac{p_j^2}{2M_j(\underline{\hat q})} + \sum_{j=\ell_0+1}^{\ell} \frac{1}{2}(p_j^2 + \omega_j^2 q_j^2) \qquad (3.1.1)$$

where $p_1, p_2, p_3, q_1, q_2, q_3$ are the momentum and position coordinates for the particles baricenter, m is their mass while $p_4, \ldots, q_4, \ldots$ are the momentum and position coordinates describing the internal degrees of freedom and $\underline{\hat q} \equiv (q_4, \ldots, q_{\ell_0})$.
Equation (3.1.1) is the form that one expects for the energy of a molecule which has a few internal degrees of freedom, precisely $\ell - \ell_0$, to which are associated oscillatory motions around equilibrium positions (corresponding to the values $j = \ell_0 + 1, \ldots, \ell$, with respective *proper frequencies* $2\pi/\omega_j$): they can therefore be called *oscillatory degrees of freedom*. The first ℓ_0 degrees of freedom describe what we shall call *translational degrees of freedom*; the position variable corresponding to a translational degree of freedom is either a position coordinate for the center of mass, varying in V, for $j = 1, 2, 3$, or an internal angular coordinate, while as a rule the variables q_j conjugated to a momentum of an oscillatory degree of freedom will always be a variable describing an internal degree of freedom and it is best thought of as varying in $(-\infty, +\infty)$.
For instance if the gas consists of point atoms with mass m then $\ell = 3$ and $E(\underline{p}, \underline{q}) = K(\underline{p}) = (p_1^2 + p_2^2 + p_3^2)/2m$. If the gas consists of diatomic

molecules built with two atoms at a fixed distance ρ, then the kinetic energy is:

3.1.2 $$K(\underline{p},\underline{q})=\frac{1}{2m}(p_1^2+p_2^2+p_3^2)+\frac{1}{2\mu}\left(\frac{p_4^2}{\rho^2}+\frac{p_5^2}{\rho^2\sin^2\theta}\right) \tag{3.1.2}$$

where m is the total mass and μ the reduced mass, $m=m_1+m_2, \mu=m_1m_2/m$, and p_4,p_5 are the momenta conjugated to the variables ϑ and φ, respectively the latitude and the azimuth of the two linked atoms. In this case the variables $\underline{q}$ conjugated to the first three momenta are real variables varying in V while the other two variables are angular variables. There are five translational degrees of freedom and no oscillatory degree of freedom.

For the perfect gases for which the total energy is a sum of the kinetic energies of the individual particles (*i.e.* the fisrt term in (3.1.2)), the partition function is:

3.1.3 $$Z(\beta,V)=\int\frac{d\underline{p}\,d\underline{q}}{h^{3N}N!}\,e^{-\beta\sum_{i=1}^N K(\underline{p}_i,\underline{q}_i)} \tag{3.1.3}$$

so that the average energy is computed by using the factorization of the integrals and by calculating explicitly first the (Gaussian) integrals over the $\underline{p}$'s and then those over the $\underline{q}$ of the oscillatory degrees of freedom and finally those over the remaining q coordinates (that are trivial if performed after those over the p's).

More generally if (3.1.1) is the kinetic energy of a single molecule and if $\underline{\hat{q}}$ are the translational position coordinates and $\underline{\tilde{q}}$ are the oscillatory ones then:

3.1.4 $$\begin{aligned}U=&\frac{\int\left(\sum_{i=1}^N K(\underline{p}_i,\underline{q}_i)\right)e^{-\beta\sum_{i=1}^N K(\underline{p}_i,\underline{q}_i)}\prod_i d\underline{p}_i\,d\underline{q}_i}{\int e^{-\beta\sum K(\underline{p}_i,\underline{q}_i)}\prod_i d\underline{p}_i\,d\underline{q}_i}=\\ =&\frac{N\int d\underline{q}(\int K(\underline{p},\underline{q})e^{-\beta K(\underline{p},\underline{q})}\,d\underline{p}\,d\underline{\tilde{q}})}{\int d\underline{\hat{q}}(\int e^{-\beta K(\underline{p},\underline{q})}\,d\underline{p}\,d\underline{\tilde{q}})}=\\ =&\frac{N\beta^{-1}\int d\underline{\hat{q}}\left(\frac{\ell_0}{2}+(\ell-\ell_0)\right)}{\int d\underline{\hat{q}}}=\\ =&N\beta^{-1}(\frac{1}{2}\ell_0+\ell-\ell_0)\,.\end{aligned} \tag{3.1.4}$$

This is an interesting relation because it is *independent* on the special form of (3.1.1) (*i.e.* independent of the coefficients $M_j(\underline{\hat{q}},\omega_i,m)$): it says that the internal energy of a perfect gas is given by the number of degrees of freedom times $1/2\beta=k_BT/2$ (equipartition of the energy among the various degrees of freedom and between kinetic and potential energy) *counting twice the oscillatory degrees of freedom* because the latter contribute to the potential energy as well. One also says the "there is equipartition between kinetic and internal elastic energy".

The constant volume specific heat of a monoatomic gas and that of a gas of *rigid* diatomic molecules are given, respectively, by:

3.1.5 $$C_V=\frac{\partial U}{\partial T}=\frac{3}{2}nR\quad\text{or}\ \frac{5}{2}nR \tag{3.1.5}$$

where $n = N/N_A$, with N_A = Avogadro's number, n the number of moles of gas and $R = k_B N_A$ is the gas constant.

Equation (3.1.5) agrees well with the experimental results on rarefied monoatomic gases, the agreement is less good for the diatomic gases, even if rarefied.

In fact (3.1.5) *cannot be accepted* on general grounds, not even for monoatomic gases (even if rarefied), because it is known that gases consist of atoms with *very many* degrees of freedom, mostly oscillatory, and their specific heat is, nevertheless $\frac{3}{2}nR$. For instance neon could be thought of as built with 20 protons and neutrons and 10 electrons, *i.e.* it would have 90 degrees of freedom, of which 87 are oscillatory(!).

But even the simple case of a truly diatomic molecule in which all internal degrees of freedom are neglected except the three describing the relative position of the two atoms is conceptually unclear: if one assumed rigidity of the distance between the two atoms then the specific heat would be $\frac{5}{2}nR$; if, instead, we admitted that the distance between the two atoms oscillates around an equilibrium position (which is more "realistic"), then the specific heat would be $\frac{7}{2}nR$ because the degrees of freedom would be 6, one of which oscillatory.

It appears, therefore, that "things go as if" some of the internal degrees of freedom were less important than others, they are "frozen" and do not contribute to the energy equipartition. Which, therefore, *would not be valid* in general, in spite of it being an extremely simple consequence of the theory of the canonical ensemble.

(II) The Specific Heat of Solids.

Another success-failure of classical statistical mechanics is the theory of the specific heat in crystalline solids. A crystalline solid can be modeled as a system of particles oscillating elastically around ideal equilibrium positions arranged on a regular lattice, *e.g.* a square lattice with mesh a (to fix ideas).

It is known from the elementary theory of oscillations that such a system is described in suitable *normal coordinates* by the Hamiltonian:

3.1.6
$$H = \sum_{\underline{k}} \frac{1}{2}(p_{\underline{k}}^2 + \omega(\underline{k})^2 q_{\underline{k}}^2) \qquad\qquad (p_{\underline{k}}, q_{\underline{k}}) \in R^6 \tag{3.1.6}$$

where the sum runs over the triples $\underline{k} = (k_1, k_2, k_3)$ of integers with $k_i = 0, 1, \ldots, \sqrt[3]{N} - 1$ if N is the number of atoms of the crystal (which we assume cubic and with side $\sqrt[3]{N} = L$), and:

3.1.7
$$\omega(\underline{k})^2 = 2c^2 \sum_{i=1}^{3} \left(1 - \cos\frac{ak_i}{L}\,2\pi\right) \tag{3.1.7}$$

with c being the sound propagation velocity in the crystal.

If we could compute the system properties by using the canonical ensemble then the internal energy could be computed as:

3.1.8
$$U = \frac{\frac{1}{2}\int\sum_{\underline{k}}\,(\underline{p}_{\underline{k}}^2 + \omega(\underline{k})^2\underline{q}_{\underline{k}}^2)\,e^{-\beta H(\underline{p},\underline{q})}\,\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}}{\int e^{-\beta H(\underline{p},\underline{q})}\,\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}} = \frac{6N}{2\beta} \qquad (3.1.8)$$

because the sum over $\underline{k}$ concerns $3N$ values and the calculation proceeds as in the case of the discussion of energy equipartition; with the difference that now *all* the $3N$ degrees of freedom are oscillatory.

Therefore the specific heat of a crystal should be:

3.1.9
$$C = 3Nk_B = 3nR \qquad (3.1.9)$$

if n is the number of moles, and this is quite well satisfied at high temperatures (above the solidification temperature, but below liquefaction) and is known as the *law of Dulong-Petit*.

If however one takes into account that a typical model of a conducting solid consists of N ions on a lattice and N electrons forming a free electron gas, then one finds that instead one should perhaps expect a specific heat of $3nR + \frac{3nR}{2}$.

Experiments show that the specific heat of crystals at high temperature indeed conforms to the Dulong-Petit law. At lower temperatures instead the specific heat approaches 0, according to a general principle called the third law of Thermodynamics.

Hence classical mechanics produces erroneous predictions also for a crystalline solid: it looks as if some degrees of freedom are *frozen* because they do not contribute to the specific heat (in other words their contribution to the internal energy is the same as that which they would give if their temperature could be considered zero and *staying* constantly so in all the system transformations: which is not possible because then the system would not be in thermal equilibrium). Furthermore at lower temperatures the crystal oscillations seem to become less and less describable by classical statistical mechanics because the specific heat deviates from the Dulong-Petit law, and tends to 0.

(III) The Black Body.

A thermodynamic theory of radiation can also be developed on the basis of the theory of ensembles, and one reaches disturbing and upsetting contradictions with the experimental observations as a consequence of classical statistical mechanics developed in the previous chapters.

In fact it was in the theory of the black body where, historically, the contradictions were felt most and led to the origin of *quantum mechanics*.

Consider a cubic region V filled with electromagnetic radiation in thermal equilibrium with the surrounding walls with which it is supposed to exchange heat. We describe the electromagnetic field by the vector potential $\underline{A}$ and the relations (which are implied by Maxwell's equations in the

vacuum):

3.1.10

$$\underline{E} = -\frac{1}{c}\frac{\partial \underline{A}}{\partial t}\,, \quad \underline{H} = \text{rot}\,\underline{A}\,, \qquad \text{div}\,\underline{A} = 0 \tag{3.1.10}$$

where c is the velocity of light, $c = 2.99 \times 10^{10}\,cm\,sec^{-1}$. It is well-known that the motion of such a field is described by the Lagrangian:

3.1.11

$$\mathcal{L} = \frac{1}{8\pi}\int_V (\underline{E}^2 - \underline{H}^2)\,\mathrm{d}\underline{x} \tag{3.1.11}$$

regarded as a function of $\underline{A}, \dot{\underline{A}}$.

If L is the side of the volume V occupied by the radiation, which for simplicity it is convenient to consider with periodic boundary conditions (by identifying the opposite sides of V), it will be possible to write $\underline{A}$ in terms of its Fourier expansion:

3.1.12

$$\underline{A}(\underline{x}) = \frac{1}{\sqrt[3]{L^3}}\sum_{\underline{k}}\sum_{\alpha=1}^{2} A^{(\alpha)}(\underline{k})\,\underline{e}^{(\alpha)}(\underline{k})\,e^{i\underline{kx}} \tag{3.1.12}$$

where $\underline{k} = \frac{2\pi}{L}\underline{\nu}$ and $\underline{\nu}$ is an integer component vector, and $\underline{e}^{(\alpha)}(\underline{k})$ are two *polarization vectors*, with unit length, and orthogonal to $\underline{k}$ and to each other. One finds

3.1.13

$$\mathcal{L} = \frac{1}{2}\sum_{\underline{k}}\sum_{\alpha=1}^{2}\left(\frac{1}{4\pi c^2}\dot{A}^{(\alpha)}(\underline{k})^2 - \frac{1}{4\pi}\underline{k}^2 A^{(\alpha)}(\underline{k})^2\right) \tag{3.1.13}$$

Therefore the evolution in time of the field in the cavity can be described by the Hamiltonian function:

3.1.14

$$\begin{aligned} H = {} & \frac{1}{2}\sum_{\underline{k}}\sum_{\alpha=1}^{2}\left(4\pi c^2 p^{(\alpha)}(\underline{k})^2 + \frac{\underline{k}^2}{4\pi} q^{(\alpha)}(\underline{k})^2\right) = \\ & = \frac{1}{2}\sum_{\underline{k}}\sum_{\alpha=1}^{2}(\,\overline{p}^{(\alpha)}(\underline{k})^2 + \underline{k}^2 c^2\,\overline{q}^{(\alpha)}(k)^2) \end{aligned} \tag{3.1.14}$$

where the pairs $(p^{(\alpha)}(\underline{k}),\ q^{(\alpha)}(\underline{k}))$ or $(\,\overline{p}^{(\alpha)}(\underline{k}), \overline{q}^{(\alpha)}(\underline{k})) = (\sqrt{4c^2\pi}p^{(\alpha)}(\underline{k}),$ $q^{(\alpha)}(\underline{k})/\sqrt{4c^2\pi})$ are canonically conjugated coordinates, equivalent because the transformation $(\underline{p}, \underline{q}) \Leftrightarrow (\,\overline{\underline{p}},\, \overline{\underline{q}})$ is canonical.

Hence an electromagnetic field in a cavity Vcan be regarded as a system of infinitely many independent harmonic oscillators.

It is, therefore, very tempting to describe this system by statistical mechanics and to say that at temperature T the microscopic states of the system will be distributed according to a canonical distribution and, hence, the probability of finding the oscillator with labels $(\alpha, \underline{k})$, *i.e.* with polarization α and wave vector $\underline{k}$, in the cell $C = dp^{(\alpha)}(\underline{k})dq^{(\alpha)}(\underline{k})$ is:

3.1.15

$$e^{-\frac{\beta}{2}(p^{(\alpha)}(\underline{k})^2 + c^2\underline{k}^2 q^{(\alpha)}(\underline{k})^2)}\,\frac{dp^{(\alpha)}(\vec{k})dq^{(\alpha)}(\vec{k})}{\sqrt{4\pi^2\beta^{-2}\underline{k}^{-2}c^{-2}}}\,. \tag{3.1.15}$$

It is clear that by assuming (3.1.15) one assumes that the cell size is negligible: this usually introduces the two types of errors that have been discussed in §2.1 and §2.4. In the present case the combinatorial error is absent because this time the oscillators are pairwise distinct. However if β is large the error due to having neglected the cell size by considering $p^{(\alpha)}(\underline{k})$ and $q^{(\alpha)}(\underline{k})$ as continuous variables is still present and it might be substantially affecting the results.

If one accepts (3.1.15) the average energy per oscillator will be $k_B T$, by the equipartition argument in (I) above, because each oscillator represents an oscillatory degree of freedom (see (3.1.14)).

Therefore it follows that, if $\nu = |\underline{k}|c/2\pi$ is the frequency of the wave with wave number $\vec{k}$, the quantity of energy $L^3 u_\nu d\nu$ corresponding to the oscillators with frequency between ν and $\nu + d\nu$ is related to the number of integer vectors $\underline{n}$ such that $\nu \le |\underline{n}|c/L \le \nu + d\nu$ via:

3.1.16
$$\begin{aligned} L^3 u_\nu d\nu =& \frac{1}{\beta} 2 \cdot (\text{number of } |\underline{n}| \text{ such that } |\underline{n}|c/L \in (\nu, \nu + d\nu)) = \\ =& \frac{2}{\beta} \big(\frac{L\nu}{c}\big)^2 \frac{4\pi L}{c}\, d\nu = L^3\, \frac{8\pi}{c^3 \beta}\, \nu^2\, \mathrm{d}\nu \end{aligned} \tag{3.1.16}$$

where in the first step the factor 2 after β^{-1} is there because, for each $\vec{k}$ there are *two* oscillators with different polarizations and equal average energy, $k_B T$. Hence the *Rayleigh-Jeans' formula* emerges:

3.1.17
$$u_\nu = \frac{8\pi\nu^2}{c^3}\, k_B T \tag{3.1.17}$$

which is manifestly in disagreement with experience, because $\int_0^\infty u_\nu\, \mathrm{d}\nu = \infty$ and a radiating cavity, in thermal equilibrium, would have infinite energy.

Experimentally the distribution (3.1.17) is observed only if ν is small, and for large ν the observations are in contrast with the energy equipartition theorem because one finds that u_ν approaches 0 very quickly as ν tends to infinity.

We see that classical statistical mechanics in the above three applications leads to paradoxes and wrong predictions. In the next section we shall see that the paradoxes disappear if the constant h measuring the cells size is $h \ne 0$: and it will be possible to say that all contradictions that appear in classical statistical mechanics arise when, to simplify the formulae by replacing summations with integrals, errors of an analytic and combinatorial nature are introduced, see also §2.1, §2.2 and §2.6, by taking $h \cong 0$.

§3.2. Classical Statistical Mechanics when Cell Sizes Are Not Negligible

In §3.1 and in the previous chapters we always neglected the size h of the phase space cells representing the microscopic states of the system. As

pointed out repeatedly, important errors are introduced in so doing (see §2.1, §2.4, which we shall see are ultimately intimately related to the paradoxes discussed in the previous sections.

The main error, however, is due to the fact that if the cells can no longer be thought of as points *then one should simply not use classical statistical mechanics*. The previous section shows that the theory leads to results that in turn permit us to *test* whether the cells in phase space can really be regarded as points: the disagreement between theory and experimental results implies that the cell size is actually accessible to experiment and this *must necessarily lead us to reformulate the very principles* of classical mechanics and therefore of statistical mechanics.

To realize how drastic could be the changes of the Thermodynamics of a system in a "quantum regime" in which h cannot be neglected one can just proceed by assuming as valid the description of the system in terms of cells in phase space and evaluate more accurately the partition sums of the various ensembles, avoiding committing the combinatorial and analytic errors that we have described above and that are really negligible only in the limit as $h \to 0$.

Consider as a first example a free gas of identical particles with no internal degrees of freedom; and let C be a generic cell of the six dimensional phase space in which the states of the single particles can be described: let the volume of C be $(\delta p\, \delta q)^3 = h^3$.

Since the identical particles are indistinguishable then the microscopic configurations Δ are determined by the numbers n_C of particles that, in the configuration Δ, occupy the cell C. Then

3.2.1
$$\begin{aligned} E(\Delta) &= \textstyle\sum_C n_C e(C) && \text{total energy} \\ N(\Delta) &= \textstyle\sum_C n_C && \text{number of particles} \end{aligned} \tag{3.2.1}$$

where $e(C)$ is the energy of a particle in the cell C.

Let us study the system in the grand canonical ensemble, where the calculations are somewhat simpler. The partition function is then:

3.2.2
$$\Xi(\beta,\lambda) = \sum_{\{n_C\}} e^{-\beta\lambda \sum_C n_C} e^{-\beta \sum_C n_C e(C)} \tag{3.2.2}$$

where, for each C, $n_C = 0, 1, 2, 3, \ldots$: see §2.5, (2.5.4).

We perform the summations explicitly, thus avoiding the combinatorial and analytical errors whose effects we are investigating. We find

3.2.3
$$\Xi(\beta,\lambda) = \prod_C \frac{1}{(1 - e^{-\beta\lambda - \beta e(C)})} \equiv e^{-\sum_C \log(1 - e^{-(\beta\lambda + \beta e(C))})} \tag{3.2.3}$$

and the probability that $n_C = n$ can be immediately computed, see (2.5.3),

3.2.4
$$p(n; C) = \frac{e^{-\beta\lambda n - \beta n\, e(C)}}{(1 - e^{-\beta\lambda - \beta\, e(C)})^{-1}} \tag{3.2.4}$$

The equation of state is deduced by expressing λ as a function of the density ρ and of β via

3.2.5
$$\rho=\rho(\lambda,\beta)=\frac{1}{V}\sum_C\sum_n n\,p(n;C)=\frac{1}{V}\sum_{n,C}\frac{n\,e^{-\beta\lambda n-\beta n e(C)}}{(1-e^{-\beta\lambda-\beta e(C)})^{-1}}\tag{3.2.5}$$

and then by replacing λ with $\lambda(\beta,\rho)$ in the grand canonical expression of the pressure. By recalling that in the grand canonical ensemble the pressure is directly related to the partition function, see (2.5.12), we get

3.2.6
$$\beta\,p(\lambda,\beta)=\frac{1}{V}\log\Xi(\beta,\lambda)=-\frac{1}{V}\sum_C\log(1-e^{-\beta\lambda-\beta e(C)})\tag{3.2.6}$$

and the total energy per unit volume u_1 is

3.2.7
$$\begin{aligned}u_1(\lambda,\beta)=&\frac{1}{V}\sum_{C,n}n\,e(C)\,\frac{e^{-\beta(\lambda+e(C))n}}{(1-e^{-\beta(\lambda+e(C))})^{-1}}=\\=&\frac{1}{V}\sum_C e(C)\frac{e^{-\beta(\lambda+e(C))}}{1-e^{-\beta(\lambda+e(C))}}\end{aligned}\tag{3.2.7}$$

To appreciate the difference between (3.2.5)-(3.2.7) and the classical perfect gas properties it is convenient to imagine that $e(C)=\underline{p}^2/2m$ if C is a cell with center at the point $(\underline{p},\underline{q})$ and, hence, to neglect the variability of $\underline{p}^2/2m$ in C.

The latter approximation implies

3.2.8
$$\begin{aligned}\beta p(\lambda,\beta)=&-\int\frac{d^3\underline{p}}{h^3}\,\log(1-e^{-\beta(\lambda+\underline{p}^2/2m)})\\\rho(\lambda,\beta)=&\int\frac{d^3\underline{p}}{h^3}\,\frac{e^{-\beta(\lambda+\underline{p}^2/2m)}}{1-e^{-\beta(\lambda+\underline{p}^2/2m)}}=\frac{1}{v}\\u_1(\lambda,\beta)=&\int\frac{d^3\underline{p}}{h^3}\,\frac{\underline{p}^2}{2m}\,\frac{e^{-\beta(\lambda+\underline{p}^2/2m)}}{1-e^{-\beta(\lambda+\underline{p}^2/2m)}}\end{aligned}\tag{3.2.8}$$

Integrating the first of (3.2.8) by parts one gets the relation:

3.2.9
$$\beta p(\lambda,\beta)=\frac{2}{3}\beta u_1\tag{3.2.9}$$

The neglect of the variability of $\underline{p}^2/2m$ in C introduces an error; however one can check, without difficulty, that it does not alter the qualitative properties of (3.2.5)-(3.2.7) which we shall discuss shortly (the approximation only simplifies the analysis, to some extent).

The most relevant phenomenon is the *Bose condensation*: the (3.2.4) show that the parameter λ must be such that $-\lambda>\min_C e(C)=0$. Hence, as

appears from (3.2.8), the maximum density $\rho_0(\beta)$ of the system seems to correspond to $\lambda = 0$:

3.2.10
$$\rho_0(\beta) = \int \frac{\mathrm{d}^3\underline{p}}{h^3} \frac{e^{-\beta\underline{p}^2/2m}}{(1 - e^{-\beta\underline{p}^2/2m})} \tag{3.2.10}$$

which looks incorrect because the density can be prescribed *a priori*, by assigning the number of particles, hence it cannot be bounded above.

But the density *can* be larger than $\rho_0(\beta)$ because (3.2.4) does not have sense if $\lambda \le 0$, $e(C) = 0$. Interpreted literally (3.2.4), for $\lambda > 0$, shows that the particle number in a cell C with $e(C) = 0$ is

3.2.11
$$\begin{aligned} \frac{\sum_{n=0}^{\infty} n e^{-\beta\lambda n}}{\sum_{n=0}^{\infty} e^{-\beta\lambda n}} &= -\frac{\mathrm{d}}{\mathrm{d}\beta\lambda} \log \sum_{n=0}^{\infty} e^{-\beta\lambda n} = \\ &= \frac{\mathrm{d}}{\mathrm{d}\beta\lambda} \log(1 - e^{-\beta\lambda}) = \\ &= \frac{e^{-\beta\lambda}}{1 - e^{-\beta\lambda}} \xrightarrow[\lambda\to 0^+]{} \infty\,. \end{aligned} \tag{3.2.11}$$

The correct interpretation of (3.2.10) and of the last remark is that the cells with $e(C) > 0$ can contribute the quantity $\rho_0(\beta)$ to the density ρ, *at most: however* the remaining part of a larger density, $\rho - \rho_0(\beta)$, is due, if $\rho > \rho_0(\beta)$, to the particles that are in the cells C with $e(C) = 0$! Note that there are *many such cells* because they must only have 0 momentum but the spatial centers of the cells can be *anywhere* in the container V.

This in fact means that the most appropriate way to describe the states of this system should be the canonical ensemble. But from the above discussion we can imagine describing a state with density $\rho > \rho_0(\beta)$ in the grand canonical ensemble by setting $\lambda = 0$ and then by imagining that $(\rho - \rho_0(\beta))V$ particles are in the cells C with $e(C) = 0$.

It is important to remark that since $\rho_0(\beta) \to 0$ for $\beta \to \infty$ the phenomenon of Bose condensation is always important at low temperature if the total density is kept fixed. And it is clear that the particles that are in the cells C with $e(C) = 0$ have zero momentum and *therefore* they do not contribute to the internal energy nor to the pressure nor to the specific heat at constant volume.

In particular if we wish to examine the specific heat at constant volume when $T \to 0$ we can note that, as soon as T is so small that $\rho_0(\beta) < \rho$ the internal energy becomes

3.2.12
$$U = u_1 V = V \int \frac{\mathrm{d}^3\underline{p}}{h^3} \frac{\underline{p}^2}{2m} \frac{e^{-\beta\underline{p}^2/2m}}{1 - e^{-\beta\underline{p}^2/2m}} = V\sigma T^{5/2} \tag{3.2.12}$$

(only cells with $e(C) \ne 0$ contribute) with

3.2.13
$$\sigma = \int \frac{\mathrm{d}^3\underline{x}}{h^3} \underline{x}^2 \frac{e^{-\underline{x}^2}}{(1 - e^{-\underline{x}^2})^{-1}} (2m)^{3/2} k_B^{5/2}\,. \tag{3.2.13}$$

Hence

3.2.14
$$C_v = \frac{\partial u_1}{\partial T} = \text{constant } T^{3/2} \quad \text{if } \rho > \rho_0(\beta) \tag{3.2.14}$$

which shows how in the perfect gas that we are studying *the equipartition result* $C_v = 3nR/2$ *is no longer true*: instead one finds $C_v \to 0$ for $T \to 0$!. At low temperatures equipartition fails if one takes $h \neq 0$ seriously.

Another example in which h cannot be neglected is the case in which the gas particles are imagined to interact in a very simple way, *conceivable although not usual* in classical mechanics: suppose that the particles "repel" each other in the sense that they cannot occupy the same cell *in phase space* so that one cannot find two or more particles in a given cell. The *unusual* nature of this force is expressed by its dependence on velocity (because it generates a "hard core" in phase space).

In the latter case the partition function is (3.2.2) with the condition that $n_C = 0, 1$. Hence:

3.2.15
$$\Xi(\lambda, \beta) = \prod_C (1 + e^{-\beta(\lambda + e(C))}) \tag{3.2.15}$$

and the probability that $n_C = n$ is, instead of (3.2.4),

3.2.16
$$p(n; C) = \frac{e^{-\beta(\lambda+e(C))n}}{1 + e^{-\beta(\lambda+e(C))}} \qquad n = 0, 1 \tag{3.2.16}$$

and (3.2.5)-(3.2.8) change accordingly.

This gas does not resemble at all the classical perfect gas and at low temperature it exhibits the phenomenon of *Fermi condensation*; one sees in fact that

3.2.17
$$p(n; C) \xrightarrow[\beta\to\infty]{} \begin{cases} 1 & \text{if } e(C) < -\lambda \\ 0 & \text{if } e(C) > -\lambda \end{cases} \tag{3.2.17}$$

so that at low temperature only the cells with $\underline{p}^2/2m < -\lambda$ are occupied: their momenta fill a sphere in momentum space (*Fermi sphere*). Note that if $\lambda > 0$ the system density tends to 0 as $T \to 0$. If one wants to keep a constant density while $T \to 0$ one must fix $\lambda < 0$. In fact if $\lambda < 0$ the density is such that:

3.2.18
$$\rho(\lambda, \beta) = \int \frac{\mathrm{d}^3\underline{p}}{h^3} \frac{e^{-\beta(\lambda+\underline{p}^2/2m)}}{1 + e^{-\beta(\lambda+\underline{p}^2/2m)}} \xrightarrow[\beta\to\infty]{} \frac{4\pi}{3h^3}\sqrt{-2m\lambda}^3 \,. \tag{3.2.18}$$

Hence if $\beta \to \infty$ and the density stays constant (*i.e.* $\lambda < 0$) one finds the internal energy and the specific heat at constant volume via the relations:

3.2.19
$$\begin{aligned} U &= V \int \frac{\underline{p}^2}{2m} \frac{\mathrm{d}^3\underline{p}}{h^3} \frac{e^{-\beta(\lambda+\underline{p}^2/2m)}}{1 + e^{-\beta(\lambda+\underline{p}^2/2m)})} \\ C_V &\equiv \left(\frac{\partial U}{\partial T}\right)_V = V \int \frac{\underline{p}^2}{2m} \frac{\mathrm{d}^3\underline{p}}{h^3} k_B \beta^2 \frac{(\lambda + \underline{p}^2/2m) e^{-\beta(\lambda+\underline{p}^2/2m)}}{(1 + e^{-\beta(\lambda+\underline{p}^2/2m)})^2} \end{aligned} \tag{3.2.19}$$

and an elementary analysis of the integrals leads to the asymptotic formula

3.2.20 $$C_V \simeq \sigma V T \qquad\qquad \text{as } T \to 0 \tag{3.2.20}$$

with a suitable σ. Hence also this system behaves in a different way if compared to the classical perfect gas at low temperature. In particular (3.2.20) shows that equipartition of energy no longer holds (because $C_V \neq V3nR/2$).

The conditions under which a behavior emerges reflecting the fact that h can no longer be considered negligible, and the classical perfect gas shows properties that are completely different from those just exhibited (without neglecting the size of h and, of course, only for small T) have been discussed in Chap.II, (2.2.2) and §2.6. We just recall that we obtained an estimate of the value of the temperature below which the effects of the nonvanishing cell size begin to be felt as:

3.2.21 $$T_q = h^2/(m k_B \rho^{-2/3})\,. \tag{3.2.21}$$

One can check, on the above formulae, that the latter value for T_q coincides, as we should expect, with the value of the temperature such that $\rho_0(\beta_q) \cong \rho$ in the first case and such that $-\lambda\beta \cong 1$ in the second.

It is common to say that the condition $T > T_q$ is the condition that the perfect gas does not present *degeneration phenomena* due to the nonnegligible size of h.

It is not difficult to realize that the degeneration due to the fact that h is appreciably $\neq 0$ can be the mechanism that permits us to avoid all the paradoxes due to energy equipartition.

For instance in the theory of a crystal, the electron contribution to the specific heat is negligible because the value of the temperature below which the electron gas presents degeneration phenomena (with consequent smallness of the specific heat, see (3.2.14) or (3.2.20)) can be estimated on the basis of (3.2.21) and it gives a very high value of T_q.

By using (3.2.21) and $m = 0.91 \times 10^{-27}\, g$, $\rho = 10^{22}\, cm^{-3}$ (density of the free electrons in iron) one finds T_q:

3.2.22 $$T_q \equiv 1/k_B\beta_q = 1.6 \times 10^5\,{}^\circ\text{K}\,. \tag{3.2.22}$$

More generally one can think that, if a given system consists of various particles, each with several internal degrees of freedom, then at a given temperature only some degrees of freedom are nondegenerate: *equipartition then takes place "between" them, while the other particles remain in a degenerate state* and therefore produce novel phenomena, among which the lack of contributions to the specific heat.

A very interesting example is that of black body radiation theory: in fact the black body is a system with infinitely many independent degrees of freedom, *most of which are in a state of extreme degeneracy* (see below),

so that the equipartition of the energy takes place *only* between a finite number of degrees of freedom.

In §3.1 we saw that a radiating cavity can be regarded as a set of infinitely many harmonic oscillators with Hamiltonian (3.1.14):

3.2.23
$$H = \frac{1}{2}\sum_{\alpha=1}^{2}\sum_{\underline{k}}(\,\overline{p}^{(\alpha)}(\underline{k})^2 + c^2\underline{k}^2\,\overline{q}^{(\alpha)}(\underline{k})^2) \tag{3.2.23}$$

where $\overline{p}^{(\alpha)}(\underline{k})$ and $\overline{q}^{(\alpha)}(\underline{k})$ are canonical variables.

The canonical distribution attributes to the configurations in which the oscillator with polarization α and wave number $\underline{k}$ is in the cell $C^{\alpha,\underline{k}}_{m,n}$ with center $(\,\overline{p}^{(\alpha)}(\underline{k}),\,\overline{q}^{(\alpha)}(\underline{k})) = (m\,\delta p, n\,\delta q)$, $(m,n$ integers), the probability:

3.2.24
$$p(C^{\alpha,\underline{k}}_{m,n}) = \frac{e^{-\frac{\beta}{2}(m^2\delta p^2 + c^2\underline{k}^2\delta q^2 n^2)}}{\sum_{\overline{m},\,\overline{n}} e^{-\frac{\beta}{2}(\,\overline{m}^2\delta p^2 + c^2\underline{k}^2\,\overline{n}^2\delta q^2)}} \tag{3.2.24}$$

where we do not neglect the dimensions of $C^{\alpha,\underline{k}}_{m,n}$, and we take seriously the canonical ensemble, forgetting that its use is doubtful when the cell sizes are not negligible and that in such cases statistical mechanics should be completely reformulated.

By repeating the analysis followed to obtain (3.2.21), see §2.2 and §2.6, one easily finds the condition under which the size of h is negligible:

3.2.25
$$\sqrt{\beta}\delta p \ll 1, \qquad \sqrt{\beta}\; c|\underline{k}|\delta q \ll 1\,. \tag{3.2.25}$$

Without explicitly fixing the values of δp and δq we see that (3.2.25) will imply, in particular (multiplying corresponding sides of the two conditions) that β is too large for a classical statistical description of the oscillators with frequency ν if

3.2.26
$$\beta\, c\,|\underline{k}|\,\delta p\,\delta q \equiv \beta h c|\underline{k}| = 2\pi\beta h\nu > 1 \tag{3.2.26}$$

where $\nu = c|\underline{k}|/2\pi$.

We must therefore expect that, given h, the high-frequency oscillators (with $|\underline{k}| \gg 1/hc\beta$ or $h\nu \gg \beta^{-1} = k_BT$) will be degenerate, *i.e.* they cannot be described without taking into account the cells sizes. Note that since ν can be as large as we want there will *always* be frequencies ν for which $h\nu \gg \beta^{-1} = k_BT$.

If we take $h = 6.62\times10^{-27}\,erg\cdot sec$ and $T = 6000\,^\circ K$ (temperature at the surface of the Sun) one sees that the degenerate frequencies are all those greater than:

3.2.27
$$\nu_0 = 1/h\beta = 1.25\times10^{14}\,cycles\,sec^{-1} \tag{3.2.27}$$

which can be compared, for the purpose of an example, with the frequency of green light (where the Sun spectrum has its maximum) $\nu_{\rm green} = 0.6\,10^{15}\,cycles\,sec^{-1}$.

The latter numerical values explain why the degeneration phenomenon has been so "easy" to observe, or "so conspicuous", in the black body radiation and why it has plaid such a big role in the development of quantum mechanics.

The average energy of a *nondegenerate* oscillator is, by energy equipartition, $k_B T = 1/\beta$, while if we do not neglect the possibility of degeneration this energy is:

3.2.28
$$u(\alpha, \underline{k}) = \sum_{m,n} \frac{1}{2}(n^2 \delta p^2 + c^2 |\underline{k}|^2 m^2 \delta q^2) p(C_{n,m}^{\alpha,\underline{k}})) \tag{3.2.28}$$

as expressed by (3.2.24).

The quantity of energy in the radiation with frequency between ν and $\nu + d\nu$ is then (see (3.1.16)):

3.2.29
$$L^3 u_\nu \, \mathrm{d}\nu = \frac{4\pi\nu^2}{c^3} \, \mathrm{d}\nu L^3 \sum_{\alpha=1}^{2} u(\underline{k}, \alpha) \tag{3.2.29}$$

where $|\underline{k}| = 2\pi\nu/c$. If $\nu \ll \frac{1}{\beta h}$ equipartition holds as one can compute explicitly using (3.2.28), (3.2.24); and (3.2.29) is simply:

3.2.30
$$u_\nu = \frac{8\pi\nu^2}{c^3} \, \beta^{-1} \, . \tag{3.2.30}$$

To discuss the high-frequency case $\nu \gg 1/\beta h$ it is necessary to fix δp and δq: *but* in classical mechanics one cannot give a clear criterion for choosing δp or δq. Hence for concreteness we shall choose δp and δq so that:

3.2.31
$$\begin{cases} \delta p \, \delta q = h \\ \delta p = \vartheta^2 c |\underline{k}| \delta q \equiv \vartheta^2 2\pi\nu \delta q \end{cases} \Rightarrow \begin{cases} \delta p = \vartheta \sqrt{2\pi\nu h} \\ \delta q = \vartheta^{-1} \sqrt{h/2\pi\nu} \end{cases} \tag{3.2.31}$$

with $\vartheta \sim 1$. Although this is a "natural" choice because it makes approximately equal the two addends in (3.2.28) for $m = n = 1$ (exactly equal if $\vartheta = 1$), it is nevertheless arbitrary. The results are qualitatively independent of the choice of ϑ, but their quantitative aspects *do depend on* its value.

From (3.2.29), (3.2.31), one deduces with a brief analysis of the series on m and n, that if $\beta h \nu \gg 1$ and $\bar{\vartheta}^2 = \min(\vartheta^2, \vartheta^{-2})$,

3.2.32
$$\begin{aligned} u(\underline{k}, \alpha) &= \tilde{h}\nu e^{-\beta \tilde{h} \nu} \qquad && \tilde{h} = 2\pi \, \bar{\vartheta}^2 \, h \quad \text{if } \vartheta \neq 1 \\ u(\underline{k}, \alpha) &= 2\tilde{h}\nu e^{-\beta \tilde{h} \nu} \qquad && \tilde{h} = 2\pi h \quad \text{if } \vartheta = 1 \end{aligned} \tag{3.2.32}$$

so that (3.2.30) yields (for $\vartheta \neq 1$ and $\tilde{h} = 2\pi h \vartheta^2$) the distribution, identical to the *Wien's distribution*,

3.2.33
$$u_\nu \cong \frac{8\pi\nu^2}{c^3} \, \tilde{h}\nu e^{-\beta \tilde{h} \nu} \tag{3.2.33}$$

which shows that the energy present at high frequency is *far* below the equipartition value and, in fact, the total energy of the electro-magnetic field in thermal equilibrium is *finite*, unlike what would happen if every oscillator had the same average energy. Of course (3.2.33) cannot really be taken seriously because, as repeatedly remarked already, the very fact that we do not neglect the size of h shows that it would be necessary to reinvestigate the basic laws of motion, based on (3.2.23), that we are using.

A further indication that (3.2.33) cannot be considered a correct distribution is seen also by noting that by changing by a small amount the cell shape (*e.g.* take $\vartheta = 1$ or $\vartheta \neq 1$ in (3.2.32)) one would find a quantitatively different result.

For instance Planck used phase space cells $C_n^{\alpha,\underline{k}}$ (for single particles) with the shape of an elliptic annulus defined by:

3.2.34
$$(n-1)h\nu \le \frac{1}{2}\,(p^{(\alpha)}(\underline{k})^2 + c^2|\underline{k}|q^{(\alpha)}(\underline{k})^2) \le nh\nu \quad n\,\text{integer} > 0 \qquad (3.2.34)$$

and area h; *i.e.* he imagined that the cells were defined by the value of the energy (and more precisely of the action to which, in this case, the energy is proportional) rather than by the momentum and position. Note that this shape is "very" different from the parallelepipedal shapes used so far.

In this way (3.2.24) and (3.2.27) are replaced by:

3.2.35
$$\begin{aligned} p(C_n^{\alpha,\underline{k}}) &= \frac{e^{-\beta h\nu n}}{(1-e^{-\beta h\nu})^{-1}} \\ u(\underline{k},r) &= \sum_{n=0}^{\infty} nh\nu\,\frac{e^{-\beta nh\nu}}{(1-e^{-\beta h\nu})^{-1}} = \frac{h\nu e^{-\beta h\nu}}{1-e^{-\beta h\nu}} \end{aligned} \qquad (3.2.35)$$

which leads to the Planck distribution:

3.2.36
$$u_\nu = \frac{8\pi\nu^2}{c^3}\,\frac{h\nu}{e^{\beta h\nu}-1} \qquad (3.2.36)$$

for the black body radiation.

Obviously on the basis of classical statistical mechanics it is impossible to decide which is the correct radiation distribution: we can only say that if in fact phase space cells cannot be chosen smaller than a minimum size, then it will be *impossible to accept equipartition* and, on the contrary, the high-frequency oscillators will have a very low average energy.

The experimental result that radiation in thermal equilibrium conforms to the Planck distribution is an indication of the non-indefinite divisibility of phase space. And the black body is a system particularly apt to reveal the discrete structure of phase space, because it consists of an infinite number of oscillators with frequency ν greater than an arbitrarily pre-fixed value ν_0, and *therefore* it contains an infinite number of degenerate oscillators if h is positive, no matter how small. In fact degeneracy happens to be already visible in "everyday life".

It is also possible, and of major interest, to investigate which would be the predictions of a strict intepretation of radiation and specific heat theories in terms of classical mechanics, *i.e.* assuming $h = 0$ in spite of our arguments in Chap.I and above, on the "unphysical nature" of such an assumption.

This is however a very difficult task and many open problems remain. Therefore I can only quote here a few papers that after the work [FPU55] have tried attacking the problem and brought a wealth of new ideas and results on the behavior of large assemblies of purely classical point particles in a situation in which the temperature is lower than the value (3.2.21) where problems with a classical interpretation begin to appear. See [GS72], [BGG84], and for the more recent developments see [BGG93], [Be94], [Be97].

§3.3. Introduction to Quantum Statistical Mechanics

In a sense quantum statistical mechanics is very similar to classical statistical mechanics: this should come as no surprise as both theories aim at explaining the same macroscopic phenomena.

As we have seen in §3.2 some of the main phenomena that receive their explanation in the framework of quantum mechanics (like the low-temperature specific heat of solids or the black body radiation or the perfect gases specific heats) can in a very qualitative and empirical sense be guessed also in classical statistical mechanics: and historically this actually happened, and sparked the genesis of quantum mechanics.

Phase space no longer has a meaning and one only thinks of observable quantities: which are described mathematically by some (few, not necessarily all, which would lead to conceptual problems, [VN55], [BH93], [Be87]) linear operators on a Hilbert space, usually infinite dimensional. On this point a long discussion could be started by arguing that this is in fact not really necessary and the dimension might be chosen finite and its size would then become a parameter that would play in quantum statistical mechanics a role similar to the phase space cells size h in classical mechanics.

However, since no "crisis" is in sight which would lead to a new mechanics, at least no crisis that is as obvious and as universally recognized as a problem like the black body radiation laws were in the early days of the twentieth century, we shall not dwell on the exercise of trying to understand how much the theory depends on the dimension of the Hilbert space or, for that matter, on the continuity of the space of the positions that particles can occupy (which one may, also, wish to challenge).

statistical ensembles are defined in terms of the *Schrödinger operator* describing the observable energy and usually denoted H. But their elements ρ, rather than as probability distributions of phase space, are defined as rules to compute the equilibrium averages of observables (which is essentially what they are used for also in classical statistical mechanics).

A ensemble $\mathcal{E}$ will, then, be a collection of rules ρ each of which allows us to compute the average value that an observable has in the macroscopic equilibrium state $\rho \in \mathcal{E}$: the element ρ should be stationary with respect to

quantum mechanical time evolution, as in the corresponding classical cases.

The mathematical notion necessary to define a "rule to evaluate averages" of observables represented by self-adjoint operators, and therefore the analog of the classical probability distributions on phase space, is that of *density matrix*. If A is an observable and H is the energy operator that corresponds to N particles in a container V, one defines the canonical ensemble as the collection of all the *density matrices* which have the form:

3.3.1
$$\rho = \, const\, e^{-\beta H} \tag{3.3.1}$$

and the average value of the observable A in the macroscopic state represented by (3.3.1), parameterized by β and V, as in the analogous case of the classical statistical mechanics, is defined by

3.3.2
$$\overline{A} = \frac{Tr A\, e^{-\beta H}}{Tr\, e^{-\beta H}} \tag{3.3.2}$$

where Tr is the *trace* operation.

As hinted above, for all practical purposes (most) operators can be regarded as big auto-adjoint matrices of large but finite dimensions, so that the trace makes sense: after some practice one in fact understands how to avoid annoying errors and pitfalls linked to this view of the operators, much in the same way in which one learns how to avoid differentiating non differentiable functions in classical mechanics.

Thermodynamics models are deduced from the (quantum) *canonical partition function*:

3.3.3
$$Z(\beta, V) = Tr\, e^{-\beta H} \tag{3.3.3}$$

and now $1/\beta k_B$ is interpreted as the temperature, while the free energy is defined by $f(\beta, v) = \lim_{V\to\infty} -\beta^{-1}\frac{1}{V}\log Z(\beta, V)$ in the limit $V \to \infty, V/N \to v$ (thermodynamic limit).

Note the the absolute temperature *is no longer defined as proportional to the average kinetic energy: rather it is identified as proportional to the parameter β^{-1} that appears in (3.3.2)*: see also §2.1, and §3.1 for a related comment on this difference. In some respects this is the really *major novelty* in quantum statistical mechanics.

One can also define quantum microcanonical or grand canonical ensembles and check their equivalence, sometimes even rigorously under suitable extra assumptions like stability and temperedness (see (2.2.17), (2.2.18)), [Ru69].

For instance, considering N identical particles of mass m in a cubic container V, the Hilbert space is the space $L_2^*(V^N)$ of symmetric (or antisymmetric) square integrable functions of the N position coordinates $(\underline{q}_1, \ldots, \underline{q}_N)$; and the energy operator is

3.3.4
$$H = -\frac{\hbar^2}{2m}\sum_{i=1}^{N} \Delta_{\underline{q}_i} + \Phi(\underline{q}) \tag{3.3.4}$$

where $\Phi(\underline{q}) = \sum_{i<j}(\underline{q}_i - \underline{q}_j)$ is the potential energy of the interaction, $\hbar = h/2\pi$ if h is Planck's constant and $\Delta_{\underline{q}_i}$ is the *Laplace operator* with respect to the i-th particle coordinate and with suitable boundary conditions (*e.g.* periodic or Dirichlet boundary conditions).

The symmetry or antisymmetry of the wave functions is imposed to take into account the specific quantum nature of the particles which can be either bosons or fermions, the latter corresponding to a system of particles which, besides the interaction energy Φ in (3.3.4), also have the extra interaction (classically nonstandard but quantum mechanically very natural) that no two particles with the same momentum can occupy the same position.

The stability notion for the interaction Φ is important in quantum statistical mechanics as much as it is important for classical statistical mechanics. An interaction is called quantum mechanically stable if there is a constant B such that the Schrödinger operator H, (3.3.4), for N identical particles satisfies, for all $N \geq 0$,

3.3.5
$$H \geq -BN \tag{3.3.5}$$

where the inequality holds in the sense of the operators (*i.e.* for any normalized quantum state $|\psi\rangle$, $\langle\psi|H|\psi\rangle \geq -BN$).

It is *interesting and important* to note that the inequality (3.3.5) can now be valid *even if* the infimum $\inf \Phi(\underline{q})$ of Φ equals $-\infty$ because the potential becomes $-\infty$ at 0 distance. In fact one can no longer separate the potential and kinetic energy as independent quantities: the *indetermination principle* in fact forbids concentrating too many particles in too small a box without giving to them a high kinetic energy. Hence there is the possibility that the decrease in potential energy due to too many close particles (contributing a large negative potential energy if $\varphi(\underline{0}) < 0$ or $\varphi(\underline{0}) = -\infty$) is compensated by the increase of kinetic energy necessary to achieve the confinement. Whether this really happens or not depends on the system (mainly on the bosonic and fermionic nature of the particles) and has to be quantitatively checked. It will be briefly discussed in Chap.IV.

The case in which the system contains several species of identical particles is treated as easily as in classical statistical mechanics. In the latter case it was sufficient to introduce suitable combinatorial coefficients to take the identity of the particles into account, see (2.2.19),(2.2.20); in the quantum case one shall simply require the symmetry or antisymmetry of the wave functions with respect to the permutations of the positions of identical particles.

For instance a system of N_1 electrically charged particles with charge $+e$ and of N_2 particles with charge $-e$ interacting with the *Coulomb force* would have, in the classical canonical ensemble, the partition function

3.3.6
$$\begin{aligned} Z(\beta,V) = \frac{1}{N_1!N_2!} \int e^{-\beta\left(\sum_{i=1}^{N_1} \underline{p}_i^2/2m_+ + \sum_{i=N_1+1}^{N_1+N_2} \underline{p}_i^2/2m_-\right)} \\ e^{-\beta\tilde{\Phi}(\underline{q})}\, \frac{\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}}{h^{3(N_1+N_2)}} \end{aligned} \tag{3.3.6}$$

with

3.3.7
$$\begin{aligned}\tilde\Phi(\underline{q})=&\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\frac{-e^2}{|\underline{q}_i-\underline{q}_{j+N_1}|}+\frac12\sum_{\substack{i,j=1\\ i\neq j}}^{N_1}\frac{e^2}{|\underline{q}_i-\underline{q}_j|}+\\ &+\sum_{\substack{i,j=1\\ i\neq j}}^{N_2}\frac{e^2}{|\underline{q}_{i+N_1}-\underline{q}_{j+N_1}|}+\sum_{i<j}^{N_1+N_2}\varphi(\underline{q}_i-\underline{q}_j)\end{aligned}\tag{3.3.7}$$

where φ is a potential describing a nonelectric force between the particles and m_+, m_- are the masses of the two species,

In the quantum case one has instead: $Z(\beta,V)=\mathrm{Tr}\,\exp-\beta H$, where H is the Schrödinger operator:

3.3.8
$$H=-\frac{\hbar^2}{2m_+}\sum_{i=1}^{N_1}\Delta_{\underline{q}_i}-\frac{\hbar^2}{2m_-}\sum_{i=1}^{N_2}\Delta_{\underline{q}_{i+N_1}}+\tilde\Phi(\underline{q})\tag{3.3.8}$$

considered as an operator acting on the space of functions $f(\underline{q}_1,\ldots,\underline{q}_{N_1},\underline{q}_{N_1+1},\ldots,\underline{q}_{N_1+N_2})$ symmetric or antisymmetric with respect to the permutations of the first N_1 variables or of the second N_2, but with no symmetry property with respect to "mixed" permutations.

The *statistics*, as one often calls the symmetry properties of the wave functions[1] with respect to the permutations of their argument plays an essential role in the theory. From the classical viewpoint adopted in §3.1 above we already had a glimpse of the phenomena that may make quantum statistical mechanics quite "strange" even from a qualitative point of view, at least at low temperatures. The reason is that the statistics can be interpreted as a special (simple) further interaction (*i.e.* no interaction in the case of bosons and a repulsive interactions *in phase space* for the fermions): compare (3.2.14) and (3.2.20).

But the statistics may play a role even at ordinary temperatures: for instance electrically neutral systems in which particles interact *only* via Coulomb forces are *unstable* in classical statistical mechanics, *at all temperatures*, for the trivial reason that the Coulomb potential between particles of opposite charge is unbounded below near the origin. But in quantum statistical mechanics they are stable if the charged particles satisfy Fermi statistics or if the bosons have charges of only one sign. See Chap.IV for a discussion of the importance of stability in statistical mechanics even in systems in which no charged particles are present.

§3.4. Philosophical Outlook on the Foundations of Statistical Mechanics

Contemporary (*i.e.* AD2000) equilibrium statistical mechanics can be said to be in an ideal conceptual stage of development.

[1] Which form the space on which the Hamiltonian acts as an operator

(1) There seem to be no fundamental theoretical problems after the irreversibility of macroscopic evolution has been shown to be compatible with the reversibility of microscopic dynamics: this understanding already came about at Boltzmann's time in terms of the existence of time scales of very different orders of magnitude over which irreversibility and reversibility can manifest themselves. It was put in a rigorous mathematical form by Lanford, see §1.8: and this should have pacified the stubborn nonbelievers in the incompatibility between microscopic reversibility and macroscopic irreversibility (it did so only very partially, in fact!).

(2) The paradoxes to which classical statistical mechanics leads have been understood in terms of *quantum effects* and the conditions of applicability of classical statistical mechanics have been correspondingly precisely formulated, see (3.2.21),(1.2.16), (2.2.2), and §2.6.

There are still many questions to be understood on the dynamics of the *approach to equilibrium* in many-particles systems and to develop reliable (and universally recognized as correct) methods to evaluate the time scales relevant in the phenomena of approach to equilibrium, see Chap.I, Chap.IX.

The ergodic problem is still not well understood particularly in systems close to mechanical equilibrium positions (as in oscillations in crystals) where it might even be conceivable that the ergodic hypothesis really fails in a substantial way, [FPU55]. This is so in spite of the major success achieved by Sinai in *proving* the ergodicity of a really interesting physical system (two balls in a periodic box, [Si70]) and its extension to many balls in a box [KSiS95]. On the whole the scarce understanding of nonequilibrium phenomena is reflected also in major problems in the kinetic theories of gases and liquids and of the related transport phenomena, [Co69],[Co93],[Do98].

The importance of the latter question has been strongly stressed by L. Galgani and by his collaborators who have devoted to the subject several important studies which led to a much better understanding of the relevance of the (probable) lack of ergodicity on time scales as long as the life of the Universe. The investigations stem from, and develop, "forgotten" remarks of "founding fathers" like Jeans, [GS72], [BGG93], [Be97].

But open problems abound also in equilibrium statistical mechanics.

The central problem of equilibrium statistical mechanics is perhaps the theory of *phase transitions* and of the corresponding *critical points*. There is no evidence of fundamental difficulties and recently some clarification has been achieved in the phase transition phenomenon as a phenomenon of *instability* with respect to boundary conditions, or of *sensitive dependence* of the equilibrium state on the boundary conditions, see Chap.II and Chap.V.

Via simple *soluble models*, see Chap.VII, it has been shown how even the simplest models of mechanical systems (like systems of magnetic spins on a lattice) can show nontrivial phase transitions and, in fact, very interesting ones. Nevertheless very important phenomena, such as the *liquid-gas transition* or the *crystal-liquid transition*, are not really understood.

In fact there is no model that could be treated avoiding approximations

which are really out of control and which describes one such transition. By "out of control" I mean approximations that have to be conceptually regarded as parts of the model itself because their influence on the results *cannot* be estimated "without hand waving". See, however, [Jo95] and the very recent [MLP98] for very encouraging steps in the right direction.

A meager consolation comes from the reassuring confirmation of the theoretical possibility (*i.e.* consistency) of such transitions accompanied by the development of many approximate theories that are continuously generated (and used in concrete applications: the ultimate goal for a wide class of scientists).

The first among such theories is the *mean field* theory which until the 1930s was the only available theory for the study of phase transitions. This is a simple theory, see Chap.V, but somewhat too rough (so as to predict phase transitions even in systems which can be shown to have none, like one-dimensional systems with short-range interactions).

The theory of phase transitions has undergone important developments mainly for what concerns the theory of *critical phenomena* in the context of which new approximate theories have been developed which provide the first real novel theoretical proposals after mean field theory; they are known as the *renormalization group* approach of Fisher, Kadanoff and Wilson, [WF72], [Wi83]. We cannot deal with such developments in this monograph: the reader will find a modern acounts of them in [BG95], [Fi98].

Another important phenomenon of equilibrium statistical mechanics is that of *metastability* and it is still not well understood: its theory involves dealing with ideas and methods (and difficulties) characteristic both of the evolution and of the equilibrium problems. Here we shall not deal with this matter, see [LP79], [CCO74], [MOS90].

Another class of not well understood phenomena are equilibrium and nonequilibrium phenomena in charged particles systems: until recently it was even qualitatively unclear how a neutral system of charged particles (*i.e. matter*) could stay in thermodynamic equilibrium, notwithstanding the strong intensity and long range of the Coulomb interaction, see [Fe85] for a detailed analysis of the basic mechanism. Until very recently only phenomenological theories for phase transitions were available, essentially based on the same type of ideas at the roots of mean field theory (for instance *Debye's screening* theory).

Recently the problem of *stability of matter* (*i.e.* of proving a lower bound proportional to N on the energy of N charged particles with zero total charge) has been satisfactorily solved in the framework of quantum statistical mechanics, [DL67], [LL72],[LT75], but the problem of a quantitative understanding of the thermodynamic equilibria in neutral aggregates of charges and of the related screening phenomena remains open, [Fe85], [Li81]. Of course there are (plenty) of very elaborate and detailed phenomenological theories, but here we mean that they are not *fundamental* and that, to be developed, require *further assumptions* (besides the fundamental assumption that equilibrium states are described, say, by the canon-

ical ensemble) that are only justified on a heuristic basis necessary to bypass otherwise non-by-passable "technical" difficulties.

For instance in the theory of molecular gases one usually postulates that a given system just consists of identical particles, with given time-invariant properties (*molecules*), that interact with each other via effective forces due to screened electromagnetic interactions. This is clearly an approximation (that we empirically think of as perfectly adequate) which obviously ignores an important part of the problem: namely that the molecules are formed by atoms which are formed by nuclei and electrons (forgetting protons, neutrons, quarks, *etc*), and the possibility that they dissociate, ionize or react chemically. Therefore one may wish to see a microscopic explanation of why in a range of densities and temperatures (absolutely crucial for our lives) matter presents itself mostly bound into complexes which have the size of isolated atoms and molecules: this is still not understood although impressive progress has been achieved in the field, [Fe85].

quantum statistical mechanics not only solves the conceptual problem of the stability of matter, [Li81], but it also introduces the possibility of a theoretical understanding of a large variety of new phenomena typically related to the quantum nature of microscopic physics: superfluidity and superconductivity are typical examples. So far such phenomena are understood only on the basis of phenomenological theories close in spirit to the mean field theory of phase transitions, [BCS57], [Br65] and Chap. 10,11 of [Fe72]. But a deeper theory has still to be developed. In fact one can say that in quantum statistical mechanics all the problems of classical statistical mechanics are present, usually in an unsolved form *even* when the corresponding classical problems are solved, and new problems that do not even exist in classical mechanics become analyzable theoretically.

It does not appear that *any* of the problems that are not understood are not understandable in the framework of statistical mechanics (classical or quantum as the case applies): no fundamental problem seems to have a theoretical description that is in conflict with experimental results. This paradisiac atmosphere may not last for long (its stability would be very surprising indeed) but as long as it lasts it gives us great peace of mind while still offering us a wide variety of fascinating unsolved problems.

Finally we mention that statistical mechanics is related to many branches of mathematics, particularly probability and information theory that have received a great influx of new ideas from the theory of phase transitions and of ensembles, [Ru69]; and the theory of dynamical systems that has received influx from the theory of approach to equilibrium, [RT71], [Ru78a], [Do98]. Combinatorics has been greatly widened by studies of the exactly soluble models in statistical mechanics, [Ba82].

Many problems in ordinary or partial differential equations have their origin in statistical mechanics which has also inspired several developments in the theory of turbulence, [Fr97], and in the theory of quantum fields, [BG95].

One can say that the present state of statistical mechanics is perhaps comparable to the state of mechanics at the moment of its triumphal applica-

tions to celestial mechanics and to ordinary mechanics at the end of the '700s and the beginning of the '800s. No obvious contradiction with experiments has yet come up and nevertheless many simple and interesting phenomena remain to be explained by the theory. A sign of vitality is also to be seen in the fact that statistical mechanics continues to generate new and deep mathematical problems: one can perhaps say, as a nontautological statement, that physical theories are sources of interesting mathematical problems only as long as they are really alive and faced with difficulties that are not purely technical.

Chapter IV:

Thermodynamic Limit and Stability

§4.1. The Meaning of the Stability Conditions

The stability and temperedness conditions (see (2.2.17),(2.2.18)):

4.1.1
$$\begin{aligned}&\Phi(\underline{q})=\sum_{i<j}\varphi(\underline{q}_i-\underline{q}_j)\ge -BN && \textit{stability}\\ &|\varphi(\underline{q}-\underline{q}')|\le C|\underline{q}-\underline{q}'|^{-3-\varepsilon}\quad\text{for}\quad |\underline{q}-\underline{q}'|\ge r_0 && \textit{temperedness}\end{aligned}\tag{4.1.1}$$

for suitable constants $C>0,\varepsilon>0,r_0>0$, have to be imposed on the interaction potential in order to insure the existence of the limits in (2.3.8), (2.3.9) or (2.5.12) defining f_c, s_m or p_{gc} respectively, *i.e.* the thermodynamic functions associated with the partition functions of various orthodic ensembles. The conditions are particularly interesting because, besides the above mathematical role, they have a simple, and profound, physical meaning.

The existence of the limits is a necessary requirement to have orthodicity and equivalence of the thermodynamic models defined by the different ensembles, as discussed in Ch.II. We have seen that if the above limits exist then the ensembles define the same model of thermodynamics for a given system. Therefore to understand the significance of the stability conditions it is convenient to examine their meaning in the thermodynamics model defined by one of the ensembles and we shall choose the canonical ensemble, where the analysis is simplest. The following analysis also illustrates some of the typical methods that are used in statistical mechanics.

(a) Coalescence Catastrophe due to Short-Distance Attraction. The first condition in (4.1.1) can be violated in several ways. One possible way is when the potential φ is negative at the origin: we always assume that φ is a smooth function for $\underline{q}\ne\underline{0}$ and, in the case at hand, we also assume that φ is smooth at the origin.

Let $\delta>0$ be fixed so small that the potential between two particles at distances $\le 2\delta$ is $\le -b<0$. Consider the canonical ensemble element with parameters β,N for a system enclosed in a (cubic) box of volume V. We want to study the probability that *all* the N particles are located in a little sphere of radius δ around the center of the box (or, for that matter, around any pre-fixed point of the box).

The potential energy of such a configuration is $\Phi\le -b\binom{N}{2}\sim -\frac{b}{2}N^2$ (because there are $\binom{N}{2}$ pairs of particles interacting with an energy $\le -b$) therefore the canonical probability of the collection $\mathcal{C}$ of such configurations will be

4.1.2
$$P_{collapse}=\frac{\int_{\mathcal{C}}\frac{d\underline{p}d\underline{q}}{h^{3N}N!}e^{-\beta(K(\underline{p})+\Phi(\underline{q}))}}{\int\frac{d\underline{p}d\underline{q}}{h^{3N}N!}e^{-\beta(K(\underline{p})+\Phi(\underline{q})}}\ge\frac{\left(\frac{4\pi}{3h^3}\right)^N\frac{\delta^{3N}}{N!}e^{\beta b\frac{1}{2}N(N-1)}}{\int\frac{d\underline{q}}{h^{3N}N!}e^{-\beta\Phi(\underline{q})}}\tag{4.1.2}$$

where we see that the contribution to the integral in the numerator can be considered as due to two factors. One is the value of the integrand function, whichwe call an "energy" factor, and one is the volume of the configuration space where the function takes the value considered, which we call an "entropy factor" or a "phase space factor". The first is β dependent while the second is not.

The phase space is extremely small because the configurations that we consider are very special: nevertheless such configurations are *far* more probable than the configurations which "look macroscopically correct", *i.e.* configurations in which the particles are more or less spaced by the average particle distance that we expect in a macroscopically homogeneous configuration namely $(\frac{N}{V})^{-1/3} = \rho^{-\frac{1}{3}}$.

The latter configurations will have a potential energy $\Phi(\underline{q})$ of the order of uN for some u, so that their probability will be bounded above by

4.1.3
$$P_{regular} \le \frac{\int \frac{d\underline{p}d\underline{q}}{h^{3N}N!} e^{-\beta(K(\underline{p})+uN)}}{\int \frac{d\underline{p}d\underline{q}}{h^{3N}N!} e^{-\beta(K(\underline{p})+\Phi(\underline{q})}} = \frac{\frac{V^N}{h^{3N}N!} e^{-\beta uN}}{\int \frac{d\underline{q}}{h^{3N}N!} e^{-\beta\Phi(\underline{q})}} \tag{4.1.3}$$

and we see that

(1) the denominators in (4.1.2),(4.1.3) are (of course) equal and

(2) the phase space factor in the numerator in (4.1.3) is much larger than the corresponding one in (4.1.2) (*i.e.* V^N against δ^{3N}), at least in the "thermodynamic limit" $V \to \infty, N \to \infty, \frac{N}{V} \to v^{-1}$.

However, no matter how small δ is, the ratio $\frac{P_{regular}}{P_{collapse}}$ will approach 0 as $V \to \infty$, $\frac{N}{V} \to v^{-1}$; extremely fast because $e^{\beta bN^2/2}$ eventually dominates over $V^N \sim e^{N \log N}$.

This means that it is far more probable to find the system in a microscopic volume of size δ rather than in a configuration in which the energy has some macroscopic value proportional to N: note that in a free gas, for instance (where $\Phi, b = 0$), the situation is the opposite, and in general, if the stability property in (4.1.1) holds, the above argument also does not apply.

This catastrophe can also be called an *ultraviolet catastrophe* as it is due to the behavior of the potential at very short distances: it causes the collapse of the system into configurations concentrated in regions as small as we please (in the thermodynamic limit).

(b) Coalescence Catastrophe due to Long-Range Attraction. This is a more interesting catastrophic behavior, because of its physical relevance. It occurs when the potential is too attractive near ∞. To simplify matters we suppose that the potential has a *hard core*, *i.e.* it is $+\infty$ for $r < r_0$, so that the above discussed coalescence cannot occur and the system cannot assume configurations in which the density is higher than a certain quantity $\rho_{cp} < \infty$, called the *close packing density.*

The catastrophe occurs if $\varphi(\underline{q}) \sim -g|\underline{q}|^{-3+\varepsilon}$, $g, \varepsilon > 0$, for $|\underline{q}|$ large. For instance this is the case of matter interacting gravitationally; if k is the gravitational constant, m is the particles mass (assuming an identical particles system), then $g = km^2$ and $\varepsilon = 2$.

In this case the probability of "regular configurations", where particles are at distances of order $\rho^{-1/3}$ from their close neighbors, is compared with that of "catastrophic configurations", with the particles at distances r_0 from their close neighbors to form a configuration in "close packing" (so that r_0 is equal to the hard core radius). Note that in the latter case the system *does not* fill the available volume and leaves empty a region whose volume is a fraction $\frac{\rho_{cp}-\rho}{\rho_{cp}} V$ of V.

A regular configuration will have a probability (in the canonical ensemble with parameters β, N and if L is the diameter of V) proportional to

4.1.4
$$P_{regular} \simeq C \frac{V^N}{h^{3N} N!} e^{g \frac{1}{2} N \rho \int_0^L |\underline{q}|^{-3+\varepsilon} d\underline{q}} \tag{4.1.4}$$

because the energy of interaction of a single particle in a medium with density ρ is, to leading order in $L \to \infty$, $\rho \int_0^L g|\underline{q}|^{-3+\varepsilon} d\underline{q} \propto \rho g L^\varepsilon$; in (4.1.4) C is a normalization constant (*i.e.* it is the reciprocal of the canonical partition function).

Likewise if we consider a configuration in close packing and we dilate it by a factor $(1+\delta)$ we obtain a configuration in which each particle can be moved in a small sphere of radius $O(\rho_{cp}^{-1/3}\delta)$ and we can call such configurations catastrophic or collapsed as they occupy only a part of the volume allowed, no matter how large the latter is (*i.e.* no matter how small ρ is compared to ρ_{cp}).

In the canonical ensemble with parameters β, N the probability of the catastrophic configurations can be bounded *below* by

4.1.5
$$P_{catastrophic} \geq C \frac{(\rho_{cp}^{-1/3}\delta)^{3N}}{h^{3N} N!} e^{g \frac{1}{2} N \rho_{cp} (1+\delta)^{-3} \int_0^L |\underline{q}|^{-3+\varepsilon} d\underline{q}} \tag{4.1.5}$$

where the constant C is the same normalization constant as in (4.1.4); and again we see that the catastrophic configurations, in spite of their very low phase space volume, have a much larger probability than the regular configurations, if $\rho < \rho_{cp}$ and δ is small enough: because the exponential in the energy factor, in (4.1.5), grows almost as $\frac{1}{2}\rho_{cp} V^{1+\varepsilon/3} \gg \frac{1}{2}\rho V^{1+\varepsilon/3}$ provided $\rho < \rho_{cp}$ and δ is small enough.

A system which is too attractive at infinity will not occupy the volume we give to it but will stay confined in a close packed configuration even in empty space.

This is important in the theory of stars: stars cannot be expected to obey "regular thermodynamics" and in particular will not "evaporate" because their particles interact via the gravitational force at large distance. Stars do not occupy the whole volume given to them (*i.e.* the universe); they do not collapse to a point only because the interaction has a strongly repulsive

core (even when they are burnt out and the radiation pressure is no longer able to keep them at a reasonable size, a reasonable size being, from an anthropocentric viewpoint, the size of the Sun).

(c) Evaporation Catastrophe: this is a another *infrared* catastrophe, *i.e.* a catastrophe due to the long-range structure of the interactions like (b) above; it occurs when the potential is *too repulsive* at ∞: *i.e.* $\varphi(\underline{q}) \sim +g|\underline{q}|^{-3+\varepsilon}$ as $\underline{q} \to \infty$ so that the temperedness condition is again violated.

Also in this case the system does not occupy the whole volume: it will generate a layer of particles *sticking* in close packed configuration to the walls of the container. Therefore if the density is lower than the close packing density, $\rho < \rho_{cp}$, the system will leave a region *around the center of the container* empty; and the volume of the empty region will still be of the order of the total volume of the box (*i.e.* its diameter will be a fraction of the box side L with the value of the fraction not depending on L as $L \to \infty$).

The proof of this statement is completely analogous to the one of the previous case, except that now the configuration with lowest energy will be the one sticking to the wall and close packed there, rather than the one close packed at the center.

Also this catastrophe is very important as it is realized in systems of charged particles bearing the *same* charge: the charges adhere to the boundary in close packing configuration and dispose themselves so that the electrostatic potential energy is minimal. We cannot, therefore, expect that the charges that we deposit on a metal will occupy the whole volume: they will rather form a surface layer minimizing the potential energy (*i.e.* so that the Coulomb potential in the interior is constant). They do not behave thermodynamically: for instance, besides not occupying the whole volume given to them, they will not contribute normally to the specific heat.

§4.2. Stability Criteria

There are simple criteria that make sure that the conditions (4.1.1) are satisfied. The first condition is satisfied, in general, if $\varphi \geq 0$: one calls this case the *repulsive potential* case, although this is a somewhat improper definition because $\varphi \geq 0$ does not imply that φ is monotonically decreasing (which would in fact generate a repulsive force in the usual sense of the word). In this case one can take $B = 0$.

Another interesting case is that of a smooth potential φ which has a non-negative Fourier transform $\hat{\varphi}$. In fact in this case:

4.2.1
$$\begin{aligned}\Phi(\underline{q}_1, \ldots, \underline{q}_n) &= -\varphi(\underline{0})n + \frac{1}{2}\sum_{i,j=1}^{n} \varphi(\underline{q}_j - \underline{q}_i) = \\ &= -\varphi(\underline{0})n + \int \hat{\varphi}(\vec{k})\frac{1}{2}\sum_{i,j=1}^{n} e^{i\vec{k}\cdot(\underline{q}_i - \underline{q}_j)} \geq -\varphi(\underline{0})n\end{aligned} \tag{4.2.1}$$

because $\sum_{i,j=1}^n e^{i\vec{k}\cdot(\underline{q}_i-\underline{q}_j)} = |\sum_{j=1}^n e^{i\vec{k}\cdot\underline{q}_j}|^2 \ge 0$. This will be called the *positive definite potential*, or *positive type*, case and one can take $B=\varphi(\underline{0})$.

Of course a potential that can be expressed as *a sum* of a positive potential and of a positive type potential is also stable. A remarkable case of a potential that can be expressed in this way is a potential such that, for $C, C', \varepsilon, r_0 > 0$, it is:

4.2.2

$$\begin{aligned}\varphi(\underline{q}) \ge& C\big(\frac{r_0}{|\underline{q}|}\big)^{3+\varepsilon}, \qquad &\text{for}\quad |\underline{q}|<r_0\\ |\varphi(\underline{q})| \le& C'\big(\frac{r_0}{|\underline{q}|}\big)^{3+\varepsilon}, \qquad &\text{for}\quad |\underline{q}|\ge r_0\,.\end{aligned} \tag{4.2.2}$$

Such a potential is sometimes called a *Lennard–Jones potential* although it is more general than the potential that was originally introduced with this name (see §1.2).

The proof of the possibility of representing φ as a sum of a positive potential and of a positive type potential can be found in [FR66]; but the stability of a potential satisfying (4.2.2) can be checked directly very simply, [Mo56]. Given, in fact, a configuration $\underline{q}_1,\ldots,\underline{q}_n$ let r be the minimum distance between pairs of distinct points: $r=\min_{i\ne j}|\underline{q}_i-\underline{q}_j|$. Suppose that the pair of closest particles is $\underline{q}_1,\underline{q}_2$ and that $r<\frac12 r_0$; then

4.2.3

$$\Phi(\underline{q}_1,\ldots,\underline{q}_n) \ge \Phi(\underline{q}_2,\ldots,\underline{q}_n)+\varphi(r)+\sum_{|\underline{q}_j-\underline{q}_1|\ge r_0}\varphi(\underline{q}_j-\underline{q}_1) \tag{4.2.3}$$

where the equality would have held had we summed over all the j's, *i.e.* also over the j's such that $|\underline{q}_1-\underline{q}_j|<r_0$.

Around each of the $\underline{q}_j$ we can draw a cube Q_j with side $\frac{r}{\sqrt{12}}$ and $\underline{q}_j$ being the vertex farthest away from $\underline{q}_1$. Since any two points among $\underline{q}_1,\ldots,\underline{q}_n$ have a distance $\ge r$ the cubes thus constructed do not overlap, and their union is contained in the complement of the sphere $|\underline{q}_1-\underline{q}|\ge r_0-\frac{r}{2}\ge\frac{r_0}{2}$; furthermore

4.2.4

$$|\underline{q}_1-\underline{q}_j|^{-(3+\varepsilon)} \le \frac{(\sqrt{12})^3}{r^3}\int_{Q_j} d\underline{q}\,|\underline{q}_1-\underline{q}|^{-(3+\varepsilon)} \tag{4.2.4}$$

so that the sum in (4.2.3) is bounded below, for some $C_1>0$, by

4.2.5

$$-C'\frac{(\sqrt{12})^3}{r^3}\int_{|\underline{q}|>\frac{r_0}{2}}\frac{r_0^{3+\varepsilon}}{|\underline{q}|^{3+\varepsilon}}d\underline{q} \ge -C_1\big(\frac{r_0}{r}\big)^3 \tag{4.2.5}$$

and $\Phi(\underline{q}_1,\ldots,\underline{q}_n)\ge C\big(\frac{r_0}{r}\big)^{3+\varepsilon}-C_1\big(\frac{r_0}{r}\big)^3+\Phi(\underline{q}_2,\ldots,\underline{q}_n)$ so that the sum of the first two terms is bounded below by some $-C_2>-\infty$, provided the assumption that the configuration $\underline{q}_1,\ldots,\underline{q}_n$ is such that $r<\frac12 r_0$ holds.

The case $r \geq \frac{1}{2}r_0$ is easier because the density is bounded above by $\sim 8r_0^{-3}$ and the interaction decreases summably at ∞: a repetition of the above considerations yields simply that $\Phi(\underline{q}_1,\ldots,\underline{q}_n) \geq \Phi(\underline{q}_2,\ldots,\underline{q}_n) + \sum_{j=2}^n \varphi(\underline{q}_1 - \underline{q}_j) \geq \Phi(\underline{q}_2,\ldots,\underline{q}_n) - C_3$ for a suitable $C_3 > 0$. Hence if $b = \max(C_2, C_3)$ then $\Phi(\underline{q}_1,\ldots,\underline{q}_n) \geq -b + \Phi(\underline{q}_2,\ldots,\underline{q}_n) \geq -bn$ and stability is proved.

We conclude by noting that (4.2.2) are sufficient stability and temperedness conditions. But in general they are far from being necessary.

In fact one would like to prove that certain systems *not fulfilling* (4.2.2) are *nevertheless* stable and have a well-defined thermodynamics, equivalently described by the canonical ensemble or by other ensembles.

For instance a gas composed of electrically charged particles and with 0 total charge is an example of a system for which we would like to prove stability, and even orthodicity of the classical ensembles.

Note that it is easy to see that a gas of charged classical particles is *not* stable: just consider the configuration in which pairs of opposite charges are put very close to each other, while the pair centers of mass are essentially equispaced; this configuration has an energy which can be made $< -bN$ for all b's, by pulling the pairs close enough together.

If, however, the particles also have a hard core interaction besides the Coulomb interaction, so that pairs of particles cannot be closer than some $r_0 > 0$, the system becomes stable as remarked by Onsager. In fact the potential of interaction between the charges will be the same as that which they would have if replaced by uniform balls of charge e_j uniformly distributed in a sphere of radius $\frac{1}{2}r_0$ around the j-th particle. Then by using the fact that there is a hard core

4.2.6
$$\Phi = \sum_{i<j}^{1,n} \frac{e_i e_j}{|\underline{q}_i - \underline{q}_j|} = \sum_{i<j}^{1,n} \int_{\{\underline{q}_i\}\times\{\underline{q}_j\}} d\underline{q}\, d\underline{q}'\, \frac{e_i \sigma_{\underline{q}_i} e_j \sigma_{\underline{q}_j}}{|\underline{q} - \underline{q}'|} \tag{4.2.6}$$

where $\sigma_{\underline{q}_i}$ does not vanish only in the sphere with radius $\frac{r_0}{2}$ around $\underline{q}_i$ and it is constant there and with integral 1; here the integration domain is the product of two balls $\{\underline{q}_i\}, \{\underline{q}_j\}$ of radius $\frac{1}{2}r_0$ centered at $\underline{q}_i$ and $\underline{q}_j$. Hence

4.2.7
$$\Phi = -\sum_{i=1}^{n} \int_{\{\underline{q}_i\}} d\underline{q}\, \frac{e_i^2 \sigma_{\underline{q}_i}^2}{|\underline{q} - \underline{q}_i|} + \frac{1}{2} \sum_{i,j}^{1,n} \int_{\{\underline{q}_i\}\times\{\underline{q}_j\}} d\underline{q} \frac{e_i \sigma_{\underline{q}_i} e_j \sigma_{\underline{q}_j}}{|\underline{q} - \underline{q}'|}\,. \tag{4.2.7}$$

If the number of species is finite it is clear that the sum in the first line can be bounded below by $-nb$, if n is the total number of particles. The double sum is simply proportional to:

4.2.8
$$\int d\vec{k}\, \frac{|\hat{\sigma}(\vec{k})|^2}{\vec{k}^2} \geq 0 \tag{4.2.8}$$

where $\hat{\sigma}(\vec{k})$ is the Fourier transform of $\sum_i e_i \sigma_{\underline{q}_i}(\underline{q})$, and we have used that the Fourier transform of the Coulomb potential is proportional to $\vec{k}^{-2}$. Hence (4.2.8) shows that a gas of a few different species of charged particles, interacting also via a hard core potential (and possibly any further additional stable potential), is stable: and one should also note that stability does not even depend on the system being neutral.

This example shows that stability can occur under situations more general than (4.2.2), and the conditions in (4.2.2) may fail and nevertheless the system may be stable.

A similar remark can be made for the temperedness condition: (4.2.2) is sufficient for temperedness but the violations of temperedness described above and leading to the above infrared catastrophes may be absent in special systems. For instance in a gas of charged particles that is *over-all neutral* this is what really happens, see [LL72].

§4.3. Thermodynamic Limit

A way to check that we are not missing some other basic condition on the potential is to show that, if the stability and temperedness conditions are satisfied, the thermodynamic limits exist and the basic ensembles are equivalent.

We have considered in Ch.II the existence of the thermodynamic limit of the entropy:

4.3.1
$$s(v,u) = \lim_{\substack{V\to\infty \\ \frac{N}{V}\to v^{-1},\frac{U}{V}\to u}} \frac{1}{Nk_B} \log \int_{H\le U} \frac{d\underline{p}\, d\underline{q}}{h^{3N} N!} \tag{4.3.1}$$

where h^{3N} is the size of a phase space cell, see Ch.I and Ch.II.

And, at least at a heuristic level, we have seen that the existence of the limit (4.3.1), the microcanonical entropy, is the key to the proof of the existence of the limits for f_c, p_{gc}, see (2.3.8),(2.5.12), and for the equivalence of the thermodynamics models based on the classical ensembles: microcanonical, canonical and grand canonical.

Therefore we shall discuss the problem of the existence of the thermodynamic limit in (4.3.1), *i.e.* the problem of the existence of the entropy in the microcanonical ensemble: this is in some sense harder than the problem of showing the existence of the corresponding limits in the canonical or grand canonical ensembles, but it has the advantage of implying the results for the other ensembles as heuristically discussed in Ch.II: the argument given there is easily turned into a proof, under very general conditions.

What follows is important not only because of the results that it establishes but also because it illustrates some of the basic techniques used in applications of statistical mechanics. In spite of its technical nature the reader could be interested in following it as it provides a good understanding of the physical meaning of the conditions (4.1.1) and of their relation with the extensivity properties of thermodynamic functions.

Before proceeding it is necessary to warn however that the following argument looks at first a bit subtle, but it is in fact quite straightforward and the intricacy is only due to the fact that we must make sure that when we divide the number of particles by certain factors (often but not always $2, 4$ or 8) we get an integer number. If one does not pay attention to this condition then the proof becomes trivial (although strictly speaking incorrect). Another source of problems is that we need to say that the partition function corresponding to a region that is the union of two subregions is the product of the partition functions corresponding to the subregions, a property that would be obvious of the regions were separated by a corridor of width larger than the interaction range. But the regions that we have to consider touch each other and we must produce corridors by shrinking the regions and therefore we must compare partition functions relative to a region and to the smaller region obtained by cutting out of it a layer of width $\frac{1}{2}r_0$ around its boundary.

The reader should, on a first reading, simply disregard the technical details related to the above counting and corridor problems and see that (4.3.6) and the consequent (4.3.8) hold at least approximately, and the existence of the limit then follows on the special sequence of cubes B_n with side $2^n L_0$ (for a fixed $L_0 > 0$ and identifying temporarily the cubes B_n and B'_n); subsequently (4.3.11), *i.e.* esentially still (4.3.6), implies both existence over arbitrary sequences of cubes and shape independence.

One could invoke the fact that the numbers of particles are so large that a change of the particle number by a few units or the taking out of a layer of width $\frac{1}{2}r_0$ around the boundary of a region that is becoming infinite, makes no difference "on physical grounds"; but *this is precisely the point*, as we must show that this is correct.

To simplify the discussion we shall suppose that the interaction has *finite range*, *i.e.* it vanishes for $|\underline{q}| > r_0$.

(A) Ground State Energy Convexity as a Function of the Energy

We first consider a special sequence of boxes and a special sequence of values of N, U. The boxes will be cubes B'_n with side size $L'_n = 2^n L_0 - r_0$ where L_0 is *an arbitrarily chosen unit of length*; the boxes B'_n are contained in the cube B_n with side size $L_n = 2^n L_0$ and stay away $\frac{r_0}{2}$, at least, from its boundaries. The volume $|B_n|$ is $2^{3n} L_0^3$.

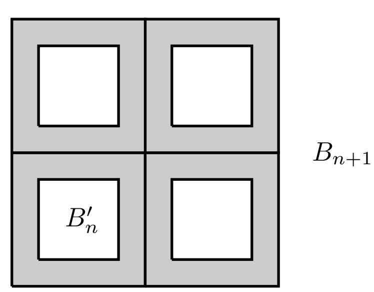


Fig. 4.3.1

The figure illustrates the boxes B_{n+1} and B'_n in the corresponding two–dimensional case (for simplicity); the region B_n is represented by the first "quarter" (*i.e.* the lower left square); the shaded areas represent the corridors of width $\frac{1}{2}r_0$ between the boundaries of the four copies of B'_n and the corresponding copies of B_n.

Given a density $\rho > 0$ we suppose first that it is *dyadic*, *i.e.* that it has the form $\rho = m2^{-3s}L_0^{-3}$ with m, s positive integers. Hence if $n \geq s$ the number $N_n = \rho|B_n|$ is an integer and we can define the "ground state energy" e_ρ at density ρ by

4.3.2
$$e_\rho = \inf_* \frac{\Phi(\underline{q}_1, \ldots, \underline{q}_{N_n})}{|B_n|} \geq -B\rho \tag{4.3.2}$$

where[1] the infimum is taken over all $n \geq s$ and over all configurations $\underline{q}_1, \ldots, \underline{q}_{N_n}$ in B'_n and the last inequality is a consequence of the stability, (4.1.1). Making a difference between B'_n and B_n is a convenient device that allows us to simplify some minor points in the forthcoming analysis and it should not be regarded as an issue. The quantity e_ρ can be thought of as the minimum energy density of the particle configurations with numerical density ρ.

Note that if ρ_1, ρ_2 are two dyadic densities then also $\rho = \frac{1}{2}(\rho_1 + \rho_2)$ is a dyadic density. Furthermore if we consider a box B'_{n+1} we see that it contains $8 = 2^3$ boxes B'_n *separated by at least* r_0.

If we put $\rho_1|B_n|$ particles in four of the boxes B'_n and $\rho_2|B_n|$ in the other 4, it is clear, from the definitions, that the energy density of any such configuration is $\leq \frac{1}{2}e_{\rho_1}|B_n| + \frac{1}{2}e_{\rho_2}|B_n|$ so that:

4.3.3
$$e_\rho \leq \frac{1}{2}(e_{\rho_1} + e_{\rho_2})\,. \tag{4.3.3}$$

In fact it is easy to see that $e_\rho \xrightarrow[\rho\to 0]{} 0$ because (4.3.2) holds together with the easily checked $e_\rho \leq \frac{1}{2}\rho \max_{|\underline{q}|>\rho^{-1}} \varphi(\underline{q})$ if ρ is small enough.[2]

Therefore if we define $e_0 = 0$ the function $\rho \to e_\rho$ is continuous on the set of dyadic densities on which it is *uniquely* defined (convex functions on the dyadics are continuous) and it can be extended by continuity to all real ρ, dyadic or not. We shall call $\underline{A}$ the *convex region* above the graph of the extension to all ρ's of $\rho \to e_\rho$.

(B) Upper Bound on Entropy.

Let $\underline{A}$ be the region $\rho > 0$ and $e > e_\rho$; let $(\rho, e) \in \underline{A}$ be an (interior) point with ρ dyadic and $e > e_\rho$. Given $\rho, e, V' \subset V$ let, to simplify the

---

[1] The quantity e_ρ is defined to be $+\infty$ if one cannot fit $\rho|B_n|$ particles inside B'_n; this may happen if the interaction contains a hard core part.

[2] This is a bound of the energy of a configuration of points regularly spaced by $\rho^{-\frac{1}{3}}$.

notations, $\mathcal{N}_0(\rho, e, V') \stackrel{def}{=} \mathcal{N}_0(U, V)$ with $\mathcal{N}_0$ the microcanonical partition function, (2.3.5), $N = \rho V$, $U = eV$. In the following V' will always be V deprived of a small corridor near its boundary.

Define, for the ρ's that are *integer multiples* of $2^{-3n}L_0^{-3}$, a corresponding *microcanonical entropy*:

4.3.4
$$\sigma_n(\rho, e) = \frac{1}{|B_n|} \log \mathcal{N}_0(\rho, e, B_n') = \frac{1}{|B_n|} \log \int_{\substack{K(\underline{p})+\Phi(\underline{q}) \le e\,|B_n| \\ \underline{q}_j \in B_n'}} \frac{d\underline{p}\, d\underline{q}}{h^{3N} N!} \tag{4.3.4}$$

Then σ_n can be bounded (because the potential Φ satisfies: $\Phi \ge -BN_n \ge -B\rho|B_n|$) by:

4.3.5
$$\begin{aligned}\sigma_n(\rho, e) &\le \frac{1}{|B_n|} \log \frac{\Omega(3N_n)\big(\sqrt{2m|B_n|(e - e_\rho)}\big)^{3N_n}}{h^{3N_n} N_n^{N_n} e^{-N_n} \sqrt{2\pi N_n}} |B_n'|^{N_n} \le \\ &\le \rho \log \frac{(4me^{5/3}/3)^{\frac{3}{2}} (e - e_\rho)^{\frac{3}{2}}}{h^3 \rho^{\frac{5}{2}}}\end{aligned} \tag{4.3.5}$$

where $\Omega(m) = 2\sqrt{\pi}^m \Gamma(\frac{m}{2})^{-1}$ is the surface of the unit sphere in m dimensional space: and no confusion should arise between the energy density e and the e (equal to the base of the natural logarithms) arising from using Stirling's formula for $N_n!$ and for the gamma function. The above inequality is an essential consequence of the stability which, in this proof, is used here for the first time.

(C) Quasi Convexity of the Entropy in Finite Volume.

The function $\sigma_n(\rho, e)$ has the simple property that it is "almost" convex. Suppose in fact that $\rho = \frac{1}{2}(\rho_1 + \rho_2)$ and $e = \frac{1}{2}(e_1 + e_2)$ with ρ, ρ_1, ρ_2 multiples of $2^{-3(n+1)}, 2^{-3n}, 2^{-3n}$, respectively, and $e_1 > e_{\rho_1}$, $e_2 > e_{\rho_2}$ (hence $e > e_\rho$). Then we can can look at the box B_{n+1}' and at 8 copies of B_n' that fit in it separated by a corridor of width r_0.

In four of the boxes B_n' we put $N_{(1)} = \rho_1|B_n|$ particles in a configuration with energy $\le e_1|B_n|$ and in the other four we put $N_{(2)} = \rho_2|B_n|$ particles in a configuration with energy $\le e_2|B_n|$; then $N = 4(N_{(1)} + N_{(2)})$. Then we shall have in B_{n+1}' exactly $N = \rho|B_{n+1}|$ particles with energy $\le e|B_{n+1}|$ and clearly

4.3.6
$$\mathcal{N}_0(\rho, e, B_{n+1}') \ge \mathcal{N}_0(\rho_1, e_1, B_n')^4 \mathcal{N}_0(\rho_2, e_2, B_n')^4 \tag{4.3.6}$$

where the $N!$ in the definition of $\mathcal{N}_0$, see (4.3.4), is *very important* because in deriving (4.3.6) one uses the fact that there are $\frac{N!}{N_{(1)}!^4 N_{(2)}!^4}$ ways of selecting which among the N particles are put in each of the eight boxes.

Relation (4.3.6) can be written in terms of the σ_n and it becomes:

4.3.7
$$\sigma_{n+1}(\rho, e) \ge \frac{1}{2}\big(\sigma_n(\rho_1, e_1) + \sigma_n(\rho_2, e_2)\big)\,. \tag{4.3.7}$$

It will be convenient to have $\sigma_n(\rho, e)$ defined not only for dyadic ρ (*i.e.* multiples of $2^{-3n}L_0^{-3}$) but for any real ρ. Since $\sigma_n(\rho, e)$ is actually defined for $\rho = \frac{m}{2^{3n}} L_0^{-3}$ this can be simply achieved by defining $\sigma_n(\rho, e)$ for the other values of ρ by *linear interpolation.*

The linear interpolation is very convenient and natural because it satisfies the bound (4.3.5), since the latter is convex in ρ, e; furthermore it has, by (4.3.7), the further property that if $\sigma_n(\rho, e)$ is *nondecreasing* in $n \ge n_0$ for each ρ a multiple of $2^{-3n_0}L_0^{-3}$, then it is *also* nondecreasing in $n \ge n_0$ for all ρ.

(D) Monotonicity of Entropy as Function of the Container Size

The property of $\sigma_n(\rho, e)$ of being nondecreasing in n, for dyadic ρ, is the key property to the analysis; in fact since, as remarked above, the box B'_{n+1} contains eight boxes B'_n separated by a corridor of width r_0 we see (from (4.3.6) with $\rho = \rho_1 = \rho_2$ and $e = e_1 = e_2$) that

4.3.8

$$\begin{aligned} \sigma_{n+1}(\rho, e) =& \frac{1}{|B_{n+1}|} \log \mathcal{N}_0(\rho, e, B'_{n+1}) \ge \\ \ge& \frac{1}{|B_{n+1}|} \log \mathcal{N}_0(\rho, e, B'_n)^8 = \\ =& \sigma_n(\rho, e) \end{aligned} \tag{4.3.8}$$

where again we make *essential* use of the $N!$ in the definition (4.3.4) of $\mathcal{N}_0$, in the same way as in the derivation of (4.3.6). By (4.3.5) $\sigma_n(e, \rho)$ is uniformly bounded in n.

Hence the limit as $n \to \infty$ of $\sigma_n(\rho, e)$ does exist for all ρ dyadic and for $e > e_\rho$. But the sequence $\sigma_n(\rho, e)$ will also converge for the nondyadic ρ's because $\sigma_n(\rho, e)$ is "almost convex" in the sense of (4.3.7), and in fact the convergence will be uniform on every closed set in the interior of $\underline{A}$. The latter property follows from elementary considerations on convex functions monotonically convergent to a limit. Clearly (4.3.7),(4.3.5) imply that the limit function $s(\rho, e)$ will be convex and bounded in the interior of the region $\underline{A}$. Hence it will be continuous in the same region.

(E) Independence from the Special Sequences of Density Values and of Growing Containers.

We now want to show that we can free ourselves from the special sequence of boxes and of densities that we have considered.

Consider a family of cubic boxes B with side $L \to \infty$ containing N particles in configurations with energy $\le E$ and suppose that $\frac{N}{V} \xrightarrow[L\to\infty]{} \rho$ and $\frac{E}{V} \xrightarrow[L\to\infty]{} e$ with (ρ, e) in the *interior* of $\underline{A}$. Below we use equivalently the notation $V = |B| = L^3$.

Let $\rho_0 = \rho + \varepsilon$ and ρ_0 be supposed dyadic, let $e_0 = e - \eta$ with ε, η small. We can divide the box B into boxes B_n of side $L_n = 2^n L_0$. Their number will be the cube of the integer part $[\frac{L}{L_n}]$ and they will cover a volume inside

B whose complement has size $\leq 8L^2L_n$. The corresponding slightly smaller boxes B_n' will therefore cover a volume

4.3.9 $$\geq L^3 - 8L^2L_n - (\frac{L}{L_n})^3 8L_n^2\frac{r_0}{2} = |B|\big(1 - 8\frac{L_n}{L} - 8\frac{r_0}{2L_n}\big)\,. \tag{4.3.9}$$

If n, L are so large that

4.3.10 $$\big(1 - 8\frac{L_n}{L} - 8\frac{r_0}{2L_n}\big)\rho_0 > \rho_0 - \frac{\varepsilon}{2} > \rho \tag{4.3.10}$$

we see that by filling each box B_n' with $N_n = \rho_0|B_n|$ particles we would put in B *more* than N particles ($N = \rho|B|$).

Hence if we fill each box B_n' with N_n particles until $[\frac{N}{N_n}]\,N_n$ particles are located and if we put in one of the remaining boxes a suitable number $< N_n$ of particles we shall have located *exactly* N particles inside B. They are put down in configurations with energy $\leq e_0|B_n|$ in each box among the first $[\frac{N}{N_n}]\,N_n$ and in a regular but arbitrary configuration in the last box, to cover as many points as necessary to reach the total correct number N (for instance on a square lattice with density higher than ρ).

Then it is clear that if L is large enough we shall have filled the box B with N particles with total energy $< E = e\,V$ and with overall density ρ; at the same time we shall have shown the inequality

4.3.11 $$\frac{1}{|B|}\log\mathcal{N}_0(\rho,e,B) \geq \frac{1}{|B|}\big(\log\mathcal{N}_0(\rho_0,e_0,B_n')^{[\frac{N}{N_n}]} + C_n\big) \tag{4.3.11}$$

where C_n bounds the $\log\mathcal{N}_0(\rho',e')$ coming from the partition function of the particles in the last box: C_n is the maximum of $\log\mathcal{N}_0(\rho',e')$ for $\rho' = \frac{m}{|B_n|}$ for $1 \leq m \leq N_n$ and e' suitably large. Once more the $N!$ in the definition of the microcanonical partition function (4.3.4) is essential.

This means, by taking into account the arbitrariness of n, that

4.3.12 $$\tilde\sigma(\rho,e) \stackrel{def}{=} \liminf_{\substack{L\to\infty\\ \frac{N}{L^3}\to\rho,\,\frac{E}{L^3}\to e}} \frac{1}{|B|}\log\mathcal{N}_0(\rho,e,B) \geq \sigma_n(\rho_0,e_0) \xrightarrow[n\to\infty]{} \sigma(\rho_0,e_0) \tag{4.3.12}$$

because $\frac{1}{|B|}C_n \xrightarrow[L\to\infty]{} 0$ and $\frac{1}{|B|}\big[\frac{N}{N_n}\big] \xrightarrow[L\to\infty]{} \frac{\rho}{\rho_n}\frac{1}{|B_n|}$.

Then the arbitrariness of e_0, ρ_0 as well as the continuity of the function $\sigma(\rho,e)$ imply that $\tilde\sigma(\rho,e) \geq \sigma(\rho,e)$.

But we can clearly repeat the argument by exchanging the role of B and B_n: namely we take a very large L so that $\frac{1}{|B|}\log\mathcal{N}_0(\rho,e,B)$ is very close to the lim sup as $L\to\infty$, $\frac{N}{L^3}\to\rho$, $\frac{E}{L^3}\to e$ of $\frac{1}{|B|}\log\mathcal{N}_0(\rho,e,B)$ and we take a very large n so that $L_n \gg L$ and show by the same type of argument that:

4.3.13 $$\sigma(\rho,e) \geq \limsup_{\substack{L\to\infty\\ \frac{N}{L^3}\to\rho,\,\frac{E}{L^3}\to e}} \frac{1}{|B|}\log\mathcal{N}_0(\rho,e,B)\,. \tag{4.3.13}$$

This will imply

4.3.14
$$\sigma(\rho,e)=\lim_{\substack{L\to\infty\\ \frac{N}{L^3}\to\rho,\,\frac{E}{L^3}\to e}}\frac{1}{|B|}\log\mathcal{N}_0(\rho,e,B)\,. \tag{4.3.14}$$

completing the proof of the existence of the microcanonical entropy over sequences of cubic boxes.

The extension to sequences of parallelepipedal boxes with all sides tending to ∞ at comparable speeds can be done along the same lines and we shall not discuss it.

(F) Box-Shape Independence.

We just mention that the argument can be perfected to show that the limit (4.3.14) exists and equals $\sigma(\rho,e)$ over much more general sequences of boxes.

Imagine, in fact, paving space with cubes of side $\ell\ge L_0$ (one of which is always centered at the origin, to fix the arbitrariness due to translation invariance); call $B_+(\ell)$ the union of the cubes of the pavement that contain points of B, and $B_-(\ell)$ the union of the cubes of the pavement entirely contained in B. Then we say that B tends to ∞ in the *sense of Fisher* if there exist $A,\alpha>0$ such that:

4.3.15
$$\frac{|B_+(\ell)|-|B_-(\ell)|}{|B|}\le A\cdot\Big(\frac{\ell}{\operatorname{diam}B}\Big)^{\alpha}. \tag{4.3.15}$$

This means that the box B grows keeping the surface small compared to the volume (homothetic growth of a box with smooth boundary trivially satisfies (4.3.15) with $\alpha=1$).

Then it can be shown that, for (ρ,e) in the interior of $\underline{A}$, (4.3.14) holds on any sequence $B\to\infty$ in the sense of Fisher.

If the potential has a hard core the same argument as above applies except that the interval of variability of ρ is no longer $[0,+\infty)$ but $[0,\rho_{cp})$, *i.e.* it is a finite interval ending at the close packing density.

The finite range condition can also be eliminated and replaced by the temperedness condition (4.3.1).

A complete treatment of all these remaining cases can be found in [Fi64], [Ru69].

The function $\sigma(\rho,e)$ is trivially related to the function $s(v,u)$ of Ch.II:

4.3.16
$$s(v,u)=v\,\sigma(v^{-1},v^{-1}u) \tag{4.3.16}$$

so that we see that $s(v,u)$ is convex in v, at fixed u, and in u at fixed v.[3]

---

[3] In fact by taking the second v–derivative of $vf(v^{-1})$ one sees that the convexity of $f(\rho)$ implies that of $vf(v^{-1})$.

This is also physically interesting as it shows that the derivative $\big(\frac{\partial s}{\partial u}\big)_v = T^{-1}$ is monotonic nonincreasing in u ("positivity of the specific heat" at constant volume) and likewise behaves $\big(\frac{\partial s}{\partial v}\big)_u = \frac{p}{T}$.

One can also see that the latter quantities are > 0, as demanded by the physical interpretation of T, p as, respectively, the absolute temperature and pressure. In fact one can prove the relations:

4.3.17
$$\begin{aligned} &\sigma_n(\rho, e+\delta) \ge \sigma_n(\rho, e) + \rho \log(1 + \frac{\delta}{e - e_\rho}) \qquad \delta > 0 \\ &\big(\frac{\partial s}{\partial v}\big)_u \xrightarrow[v\to\infty]{} 0^+ \end{aligned} \tag{4.3.17}$$

which together with the convexity properties immediately imply the positivity of T, p.

(G) Continuity of the Pressure

Consider the pressure as a function of the density at constant temperature. This is a function that is most conveniently studied in the grand canonical ensemble. Since pressure is a convex function of the chemical potential[4] λ its derivative is monotonically nondecreasing (because convexity means that the first derivative is nondecreasing); hence its graph has, at most, countably many upward jumps and countably many horizontal plateaus. A vertical jump corresponds to a value of the chemical potential where the right and left derivatives of the pressure are different, while a horizontal plateau corresponds to a straight segment in the graph of the pressure (drawing a few schematic graphs is very helpful here).

The definition of grand canononical partition function implies, by differentiating it with respect to λ, that its λ–derivative is the density. Then to say that there are no horizontal plateaus in the graph of the density as a function of the chemical potential is equivalent to saying that the graph of the pressure as a function of the chemical potential contains no straight segment.

A horizontal plateau in the graph of the density as a function of the chemical potential means that there are several chemical potentials corresponding to the same density.[5] Hence there are several pressures corresponding to

[4] Because $\beta p = \frac{1}{V} \log \Xi(\beta, \lambda)$ and $\Xi(\beta, \lambda)$, the grand canonical partition function, see Ch.II, is $\Xi = \sum_{n=0}^{\infty} \frac{e^{-\beta\lambda\, n}}{n!} Z'_n(\beta)$ with Z'_n the canonical partition function for n particles; hence *in general* the logarithm of a sum of quantities $ce^{c'\lambda}$ with $c > 0$ is a convex function of λ as one checks that the second derivative of this sum with respect to λ is nonnegative.

[5] Note that the derivative of the pressure with respect to the chemical potential is the density and, therefore, it *should* be strictly positive so that the pressure is strictly increasing with the chemical potential. But one has to show that the density is positive if the chemical potential is negative enough: this is not so easy and it will be shown in complete generality in §5.9: see the sentence preceding (5.9.17) where the relation $\rho = e^{\lambda\beta}(1 + O(e^{\lambda\beta}))$ is given and interpreted as saying that the density is proportional to the activity $z = e^{\beta\lambda}$ at small activity *i.e.* at negative enough chemical potential.

the same density, *i.e.* a vertical *discontinuity* in the graph of the pressure as a function of the density.

It is therefore interesting to see that such discontinuities of the pressure as a function of the density at constant T are in fact not possible under the only condition that the potential is superstable; see Ch.V, (5.3.1) for a definition.

A simple proof, due to Ginibre, [Gi67], can be given for systems with hard core or, alternatively, with purely repulsive pair potential. The general case of a supertsable potential is harder, [DM67],[Ru70].

Define the *activity* as $z \stackrel{def}{=} e^{-\beta\lambda}\sqrt{\pi\, 2m\beta^{-1}}^3 h^{-3}$, where the second factor comes from integrating explicitly the kinetic part of the energy so that the grand canonical partition function Ξ (in volume V and temperature $T=(k_B\beta)^{-1}$) can eventually be written

$$\Xi = \sum_{n=0}^{\infty} \frac{z^n}{n!} Z_n,$$

see below. Then we estimate

$$\chi^{-1} = \frac{\rho}{\beta}\frac{\partial\beta p}{\partial\rho} \equiv \frac{\rho}{\beta}\frac{\partial\beta p}{\partial\beta\lambda}\frac{\partial\beta\lambda}{\partial\rho} \equiv \frac{\rho}{\beta} z \frac{\partial\beta p}{\partial z}\,\frac{1}{z}\frac{\partial z}{\partial\rho}$$

at constant β: in the grand canonical ensemble it is $\beta p = \frac{1}{V}\log\Xi$. The physical interpretation of χ is clearly that of *isothermal compressibility*.

If we can show that χ^{-1} can be bounded away from $+\infty$ for $z = e^{\beta\lambda}$ in any finite interval then χ is bounded away from zero in any finite interval and the graph of the pressure as a function of the density is continuous (in fact Lipshitz continuous) and as a function of the chemical potential it cannot contain any straight segment.

The grand canonical density is $\rho = \frac{\langle n\rangle}{V}$ with $\langle n\rangle \stackrel{def}{=} \sum_{n=0}^{\infty} n \frac{z^n}{n!}\frac{Z_n}{\Xi}$, where

4.3.18 $$Z_n = \int_{V^n} dx_1 \ldots dx_n\, e^{-\beta \sum_{i<j}\varphi(x_i - x_j)} \tag{4.3.18}$$

with $\Xi = \sum_{n=0}^{\infty} \frac{1}{n!} z^n Z_n$. Hence, if $\beta p = V^{-1}\log\Xi$, then:

$$\rho = z\frac{\partial\beta p}{\partial z} = \frac{\langle n\rangle}{V} \text{ and } z\frac{\partial\langle n\rangle}{\partial z} = \langle n^2\rangle - \langle n\rangle^2;$$

furthermore

$$\chi^{-1} = \frac{\rho}{\beta}\frac{\partial\beta p}{\partial\rho} = \frac{\rho}{\beta}\frac{\partial\beta p}{\partial z}(\frac{\partial\rho}{\partial z})^{-1}.$$

Hence:

4.3.19 $$\chi^{-1} = \frac{\rho}{\beta} z\frac{\partial\beta p}{\partial z}(z\frac{\partial\rho}{\partial z})^{-1} = \frac{\rho}{\beta}\frac{\langle n\rangle}{\langle n^2\rangle - \langle n\rangle^2} \tag{4.3.19}$$

and we need a lower bound on $\frac{\langle n^2\rangle - \langle n\rangle^2}{\langle n\rangle}$.

Consider systems *with hard core interactions*, *i.e.* with a pair potential $\varphi(\underline{x}-\underline{y})$ equal to $+\infty$ for $|\underline{x}-\underline{y}|\le a$ for some $a>0$, or systems with repulsive potential $\varphi\ge 0$. Let $\varphi_+(\underline{q})\stackrel{def}{=}\varphi(\underline{q})$ if $\varphi(\underline{q})\ge 0$ and $\varphi_+(\underline{q})\stackrel{def}{=}0$ otherwise.

The key remark is that $Z_{n+2}/Z_{n+1}\ge Z_{n+1}/Z_n-D$ for some *volume independent and continuous* function $D=D(\beta)$. Hence in the hard core case one can take $D=e^{\beta B}\int(1-e^{-\beta\varphi_+(x)})\,dx<+\infty$ if $-B$ is a lower bound for the energy of interaction Φ_1 of one particle with any number of others.[6]

Setting $\Phi(X)$ the potential energy of the configuration X this follows from the Schwarz inequality; abbreviating $(\underline{x}_1,\ldots,\underline{x}_n)\stackrel{def}{=}X$ and $\Phi_1(X,\underline{x})\stackrel{def}{=}\sum_{j=1}^n\varphi(\underline{x}_j-\underline{x})\ge -B$, we find:

4.3.20
$$\begin{aligned}
Z^2_{n+1}\equiv&\Big(\int dXe^{-\beta\Phi(X)}\int d\underline{x}e^{-\beta\sum_{j=1}^n\varphi(\underline{x}_j-\underline{x})}\Big)^2\le\\
&\le\Big(\int dXe^{-\beta\Phi(X)}\Big)\cdot\Big(\int dX\,d\underline{x}\,d\underline{y}e^{-\beta\Phi(X)-\beta\Phi_1(X,\underline{x})-\beta\Phi_1(X,\underline{y})}\Big)\equiv\\
&\equiv Z_n\int dXd\underline{x}d\underline{y}e^{-\beta\Phi(X,\underline{x},\underline{y})}(e^{+\beta\varphi(\underline{x}-\underline{y})}-1+1)=\\
&=Z_nZ_{n+2}+Z_n\int dXd\underline{x}e^{-\beta\Phi(X,\underline{x})}\int d\underline{y}e^{-\beta\Phi_1(X,\underline{y})}(1-e^{-\beta\varphi(\underline{x}-\underline{y})})\le\\
&\le Z_nZ_{n+2}+Z_n\int dXd\underline{x}e^{-\beta\Phi(X,\underline{x})}\int d\underline{y}e^{\beta B}(1-e^{-\beta\varphi_+(\underline{x}-\underline{y})})\equiv\\
&\equiv Z_nZ_{n+2}+Z_nZ_{n+1}D,\quad\longleftrightarrow\quad\frac{Z_{n+2}}{Z_{n+1}}\ge\frac{Z_{n+1}}{Z_n}-D
\end{aligned}\tag{4.3.20}$$

where $D=e^{\beta B}\int(1-e^{-\beta\varphi_+(\underline{y})})d\underline{y}$. Therefore, using again the Schwartz inequality (plus the normalization property $\sum_n\frac{z^nZ_n}{n!\Xi}\equiv 1$)

4.3.21
$$\begin{aligned}
(\langle n\rangle(1+zD))^2\equiv&\Big(\sum_{n=0}^\infty\frac1{n!}z^n\frac{Z_n}{\Xi}\big(\frac{z^{n+1}Z_{n+1}}{z^nZ_n}+nzD\big)\Big)^2\le\\
&\le\sum_{n=0}^\infty\frac1{n!}z^n\frac{Z_n}{\Xi}\big(\frac{z^{n+1}Z_{n+1}}{z^nZ_n}+nzD\big)^2
\end{aligned}\tag{4.3.21}$$

so that developing the square the r.h.s. becomes

$$\sum_{n=0}^\infty\frac{z^n}{n!}\frac1\Xi\Big(z^2\frac{Z^2_{n+1}}{Z_n}+2z^2D\,n\,Z_{n+1}+z^2\,n^2\,D^2\,Z_n\Big)$$

[6] Therefore $B<+\infty$ if there is a hard core or if $\varphi\ge 0$; while if the potential is of Lennard–Jones type, or more generally a superstable potential, with an attractive part there is no such B and the theory is more difficult, see [Ru70].

and using the last of (4.3.20) this is bounded above, if $p_n = (z^n/n!)\,(Z_n/\Xi)$, by

$$\begin{aligned}
&\sum_{n=0}^{\infty} \frac{z^n}{n!}\frac{1}{\Xi}\big(z^2 Z_{n+2} + z^2 D Z_{n+1} + 2z^2 D n Z_{n+1} + n^2 z^2 D^2 Z_n\big) = \\
&= \sum_{n=0}^{\infty} \big((n+1)(n+2)p_{n+2} + zD\,(n+1)p_{n+1} + \\
&+ 2zD\,n(n+1)p_{n+1} + z^2 D^2\, n^2 p_n\big) = \\
&= \langle n^2\rangle - \langle n\rangle + zD\langle n\rangle + 2zD(\langle n^2\rangle - \langle n\rangle) + z^2 D^2 \langle n^2\rangle = \\
&= (1+zD)^2\langle n^2\rangle - (1+zD)\langle n\rangle
\end{aligned}$$

hence one finds

4.3.22
$$\frac{\langle n^2\rangle - \langle n\rangle^2}{\langle n\rangle} \geq \frac{1}{1+zD} \tag{4.3.22}$$

which proves a lower bound on χ and, therefore, the continuity of the pressure as a function of the density at constant temperature.

.

Chapter V:

Phase Transitions

§5.1. Virial Theorem, Virial Series and van der Waals Equation

Van der Waals theory is one of the earliest and simplest applications of classical statistical mechanics (1873: see [VW88]). Nevertheless it brings up one more of its conceptual problems, although not as deep as the critical problems of Chap.III. It clearly indicates that one has to give up the naive hope that the theory of phase transitions and phase coexistence could be easily quantitatively accessible.

The classical approach starts from the *virial theorem* (Clausius). Consider a real gas with N identical particles with mass m in a spherical (for simplicity) container with volume V; suppose that the microscopic interaction potential between two particles at distance r is a Lennard-Jones potential:

5.1.1
$$\varphi(r) = 4\varepsilon((r_0/r)^{12} - (r_0/r)^6) \tag{5.1.1}$$

where ε is the interaction *strength* and r_0 is the *diameter* of the molecules.

Let the force acting on the i-th particle be $\underline{f}_i$; multiplying both sides of the equations of motion $m\ddot{\underline{q}}_i = \underline{f}_i$ by $-\frac{1}{2}\underline{q}_i$ we find

5.1.2
$$-\frac{1}{2}\sum_{i=1}^{N} m\underline{q}_i \cdot \ddot{\underline{q}}_i = -\frac{1}{2}\sum_{i=1}^{N} \underline{q}_i \cdot \underline{f}_i \stackrel{def}{=} \frac{1}{2}C(\underline{q}) \tag{5.1.2}$$

and the quantity $C(\underline{q})$ is the *virial* of the forces in the configuration $\underline{q}$; note that $C(\underline{q})$ is *not* translation invariant because of the presence of the forces due to the walls: writing the force $\underline{f}_i$ as a sum of the internal forces and of the external forces, due to the walls, the virial C can be expressed naturally as sum of the virial C_{int} of the internal forces (translation invariant) and of the virial C_{ext} of the external forces. By dividing both sides by τ and integrating over the time interval $[0,\tau]$ one finds, in the limit $\tau \to +\infty$,

5.1.3
$$\langle T\rangle = \frac{1}{2}\langle C\rangle \tag{5.1.3}$$

which is read by saying that the average kinetic energy equals half the average virial of the forces.

The virial naturally splits as the sum of the virial due to the internal forces C_{int} and that due to the external ones C_{ext}. The virial of the external forces is simply

5.1.4
$$\langle C_{ext}\rangle = 3pV \tag{5.1.4}$$

where p is the pressure and V the volume. Equations (5.1.3) and (5.1.4) constitute the *virial theorem* of Clausius.

A quick proof of (5.1.4) is that the external forces act only on the boundary of the (spherical) box B containing the system: they send back into the

container particles that try to get out. The average force that a surface element $d\sigma$ of the walls exercises over each colliding particle is $-p\underline{n}_e\,d\sigma$ if p is the pressure and $\underline{n}_e$ the outer normal. Hence $\langle C_{ext}\rangle = p\int_{\partial B} d\sigma_\xi \underline{\xi}\cdot\underline{n}_e(\xi) = p3V$ by Green's volume formula. A more refined argument leading to the same result is possible.[1]

Since the average kinetic energy is $\frac{3}{2}\beta^{-1}N$ we see that the equation of state is:

5.1.5
$$\frac{1}{\beta} = pv + \frac{\langle C_{int}\rangle}{3N} \tag{5.1.5}$$

if $\beta^{-1} = k_B T$ and T is the absolute temperature and v the specific volume.

The two relations (5.1.3) and (5.1.4) together with their corollary (5.1.5) constitute *Clausius' virial theorem.* The equation (5.1.5) is essentially the equation of state. In the case of no internal forces it yields $\beta pv = 1$, the ideal gas equation.

Van der Waals first used the virial theorem to perform an actual computation of the corrections. Note that the internal virial C_{int} can be written, if $\underline{f}_{j\to i} = -\underline{\partial}_{\underline{q}_i}\varphi(\underline{q}_i - \underline{q}_j)$

5.1.6
$$C_{int} = -\sum_{i=1}^{N}\sum_{i\neq j}\underline{f}_{j\to i}\cdot\underline{q}_i \equiv \sum_{i<j}\underline{\partial}_{q_i}\varphi(\underline{q}_i - \underline{q}_j)\,\cdot\,(\underline{q}_i - \underline{q}_j) \tag{5.1.6}$$

which shows that the contribution to the virial by the internal repulsive forces is *negative* while that of the attractive forces is *positive.* To evaluate the average of (5.1.6) we simply use the theory of the ensembles and choose to use the canonical ensemble, as it is more convenient.

---

[1] The force due the spherical container boundary can be represented as:

$$\underline{\delta}(\underline{x}) = -\int_{\partial B} d\sigma_{\underline{\xi}}\,\underline{n}^e(\underline{\xi})F(\underline{x}-\underline{\xi}) \tag{a}$$

where F is a nonnegative scalar, $\underline{n}^e(\underline{\xi})$ is the outer normal to the boundary ∂B of B at $\underline{\xi}$. The function F is not zero and very intense only in a very tiny region near the origin, so that $\underline{\delta}(\underline{x})$ is not zero only very close to the boundary. We are really interested in the limiting case in which the force F is a Dirac δ-function, which represents the ideal case of a perfect wall with no width.

The virial of the external forces $\underline{\delta}$ necessary to confine the system inside the box B is

$$-\langle\sum_i \underline{x}_i\cdot\underline{\delta}(\underline{x}_i)\rangle = \int_{\partial B} d\sigma_{\underline{\xi}}\underline{n}^e(\underline{\xi})\cdot\underline{\xi}\,\langle\sum_i F(\underline{x}_i - \underline{\xi})\rangle \tag{b}$$

where $\langle\cdot\rangle$ denotes the time avergae, and having replaced $\underline{n}^e(\underline{\xi})\cdot\underline{x}_i$ by $\underline{n}^e(\underline{\xi})\cdot\underline{\xi}$ because of the locality property of F (exact if δ is a delta function). The average $\langle\sum_i F(\underline{x}_i - \underline{\xi})\rangle$ is $\underline{\xi}$ independent because of the assumed spherical symmetry and it represents the force exerted per unit surface area near $\underline{\xi}$, *i.e.* it is the pressure p so that the average virial is $p\int_{\partial B} d\sigma_{\underline{\xi}}\underline{n}^e_{\underline{\xi}}\cdot\underline{\xi} = 3pV$. For an extension to nonspherical containers see [MP72].

One could proceed by using time averages and the ergodic hypothesis, *i.e.* the microcanonical ensemble, but the result would be the same. One could also proceed by simply taking the few time averages that we really need and argue that their value should coincide with the one we calculate assuming the gas essentially free and adding corrections that take into account that close particles interact. This would be far weaker than the ergodic hypothesis and it was the path followed by van der Waals; however, unlike the equidistribution assumption, it does not lead easily to a systematic power series expansion in v^{-1} of the corrections.

In the canonical ensemble the average internal virial is, taking into account the symmetry in $\underline{q}_i$ and denoting $\Phi_{q_1}(\underline{q}_3,\ldots,\underline{q}_N)=\sum_{j=3}^{N}\varphi(\underline{q}_1-\underline{q}_j)$:

5.1.7

$$\begin{aligned}\langle C_{int}\rangle = \binom{N}{2}\int \frac{d\underline{q}_1 d\underline{q}_2\cdots}{N!} e^{-\beta\Phi_{\underline{q}_1}(\underline{q}_3,\ldots,\underline{q}_N)-\beta\Phi(\underline{q}_2,\ldots,\underline{q}_N)}\cdot\\ \cdot e^{-\beta\varphi(\underline{q}_1-\underline{q}_2)}\,\underline{\partial}_{\underline{q}_1}\varphi(\underline{q}_1-\underline{q}_2)\cdot(\underline{q}_1-\underline{q}_2)\end{aligned}\tag{5.1.7}$$

which can be rewritten (using an integration by parts) as

5.1.8

$$\begin{aligned}\langle C_{int}\rangle = &\binom{N}{2}\int \frac{d\underline{q}_1 d\underline{q}_2\cdots}{N!} e^{-\beta\Phi_{\underline{q}_1}(\underline{q}_3,\ldots,\underline{q}_N)-\beta\Phi(\underline{q}_2,\ldots,\underline{q}_N)}\cdot\\ &\cdot\frac{-1}{\beta}\,\underline{\partial}_{\underline{q}_1}\big(e^{-\beta\varphi(\underline{q}_1-\underline{q}_2)}-1\big)\ \cdot(\underline{q}_1-\underline{q}_2)=\int\frac{d\underline{q}_1 d\underline{q}_2}{2}\,\frac{3}{\beta}\cdot\\ &\cdot\int\frac{d\underline{q}_3\cdots}{(N-2)!}\big(e^{-\beta\varphi(\underline{q}_1-\underline{q}_2)}-1\big)e^{-\beta\Phi_{\underline{q}_1}(\underline{q}_3,\ldots,\underline{q}_N)-\beta\Phi(\underline{q}_2,\ldots,\underline{q}_N)}-\\ &-\int\frac{d\underline{q}_1 d\underline{q}_2 d\underline{q}_3}{2}\int\frac{d\underline{q}_4\ldots d\underline{q}_N}{(N-3)!}\cdot e^{-\beta\Phi_{\underline{q}_1}(\underline{q}_3,\ldots,\underline{q}_N)-\beta\Phi(\underline{q}_2,\ldots,\underline{q}_N)}\cdot\\ &\cdot\big(e^{-\beta\varphi(\underline{q}_1-\underline{q}_2)}-1\big)(\underline{q}_1-\underline{q}_2)\cdot\underline{\partial}_{\underline{q}_1}\varphi(\underline{q}_1-\underline{q}_3)\end{aligned}\tag{5.1.8}$$

where the $\partial_{q_1}e^{-\beta\varphi(\underline{q}_1-\underline{q}_2)}$ has been replaced, before integrating by parts, by $\partial_{q_1}\big(e^{-\beta\varphi(\underline{q}_1-\underline{q}_2)}-1\big)$ to avoid boundary contributions in the integrations (in fact $e^{-\beta\varphi(\underline{q})}$ is 1 at $\underline{q}=\infty$ when, eventually, we take the limit as $V\to\infty$).

To rewrite (5.1.8) in a better form it is useful to introduce the notion of *correlation function*: the k-points correlation function $\rho(\underline{q}_1,\ldots,\underline{q}_k)$ is defined so that $\rho(\underline{q}_1,\ldots,\underline{q}_k)\frac{d\underline{q}_1\cdots\underline{q}_k}{k!}$ is the probability of finding k particles in the infinetisimal volume elements $d\underline{q}_1,\ldots,d\underline{q}_k$. Hence, in the canonical ensemble:

5.1.9

$$\rho(\underline{q}_1,\ldots,\underline{q}_k)=Z_N^{-1}\int\frac{d\underline{x}_1\ldots d\underline{x}_{N-k}}{(N-k)!}e^{-\beta\Phi(\underline{q}_1,\ldots,\underline{q}_k,\underline{x}_1,\ldots,\underline{x}_{N-k})}\tag{5.1.9}$$

where the normalization Z_N is the canonical partition function. Note that $\rho(\underline{q}_1,\ldots,\underline{q}_k)$ is *not* normalized to 1; in fact $\int_V\rho(\underline{q}_1,\ldots,\underline{q}_k)d\underline{q}_1\ldots d\underline{q}_k=$

$N(N-1)\dots(N-k+1)$. It is a simple matter of algebra to check that

5.1.10
$$\begin{aligned}\langle C_{int}\rangle = \frac{3}{2\beta}\int d\underline{q}_1 d\underline{q}_2 \big(e^{-\beta\varphi(\underline{q}_1-\underline{q}_2)}-1\big)e^{\beta\varphi(\underline{q}_1-\underline{q}_2)}\rho(\underline{q}_1,\underline{q}_2)+\\ +\frac{1}{2\beta}\int d\underline{q}_1 d\underline{q}_2 d\underline{q}_3 \big(e^{-\beta\varphi(\underline{q}_1-\underline{q}_2)}-1\big)(-\beta)\cdot\\ \cdot\underline{\partial}_{\underline{q}_1}\varphi(\underline{q}_1-\underline{q}_3)\cdot(\underline{q}_1-\underline{q}_2)e^{\beta\varphi(\underline{q}_1-\underline{q}_2)}\rho(\underline{q}_1,\underline{q}_2,\underline{q}_3)\,.\end{aligned} \tag{5.1.10}$$

It can be shown, see §5.9 below, that the correlation functions of order k are, for ρ small, analytic functions of ρ divisible by ρ^k and proportional to $e^{-\beta\Phi(\underline{q}_1,\dots,\underline{q}_k)}$; in fact

5.1.11
$$\rho(\underline{q}_1,\dots,\underline{q}_k)=\rho^k\, e^{-\beta\Phi(\underline{q}_1,\dots,\underline{q}_k)}\,(1+\rho F_1(\underline{q})+\rho^2F_2(\underline{q})+\dots) \tag{5.1.11}$$

so that (5.1.10) can be used as a starting point for a systematic expansion of the equation of state in powers of ρ. We do not discuss here the possibility of the expansion (5.1.11), not because it is difficult, but because it would lead us into a technical question that certainly was not worrying people at the time the above analysis was performed; we defer it to §5.9.

The physical meaning of the correlation functions of order k shows that they should be proportional to ρ^k and their definition (5.1.9) shows that they ought also to be proportional to $e^{-\beta\Phi(\underline{q}_1,\dots,\underline{q}_k)}$. Hence it is quite clear that unless some integrals diverge, (5.1.10) already allows us to evaluate the first correction to the gas law. We simply neglect the third order term in the density and use $\rho(\underline{q}_1,\underline{q}_2)=\rho^2e^{-\beta\varphi(\underline{q}_1-\underline{q}_2)}$ in the second order term.

But there is no apparent reason for the integrals to diverge: they contain the factors $\big(e^{-\beta\varphi(\underline{q}_1-\underline{q}_2)}-1\big)$ and $\partial_{\underline{q}_1}\varphi(\underline{q}_1-\underline{q}_3)$ which tend to zero at large arguments so that the divergence sources should be quite subtle. About hundred years after the original work of van der Waals the actual convergence of the series in (5.1.11) and of the virial series has been mathematically proved. We shall discuss it from a modern viewpoint in the following §5.9.

Then

5.1.12
$$\frac{1}{2}\langle C_{int}\rangle = V\frac{3}{2\beta}\rho^2 I(\beta)+VO(\rho^3) \tag{5.1.12}$$

where $I(\beta)=\frac{1}{2}\int\big(e^{-\beta\varphi(q)}-1\big)\,d^3q$ and the equation of state (5.1.5) becomes $pv+\frac{I(\beta)}{\beta v}+O(v^{-2})=\beta^{-1}$.

The calculation of I can be performed approximately if $\beta\varepsilon\ll 1$ (*i.e.* at "high temperature"), by imagining that $\varphi(r)=+\infty$ (*i.e.* $e^{-\beta\varphi(r)}-1=-1$) for $r<r_0$ and $e^{-\beta\varphi(r)}-1=-\beta\varphi(r)$ for $r>r_0$. One has:

5.1.13
$$\begin{aligned}I &\cong \frac{1}{2}\int_0^{r_0}-4\pi r^2dr-\frac{\beta}{2}\int_{r_0}^{\infty}\varphi(r)4\pi r^2dr=\\ &=-4v_0+\frac{32}{3}\beta\varepsilon v_0=\\ &=-(b-\beta a)\end{aligned} \tag{5.1.13}$$

with

5.1.14
$$v_0 = \frac{4\pi}{3}(\frac{r_0}{2})^3, \quad b = 4v_0, \quad a = \frac{32}{3}\varepsilon v_0 \,. \tag{5.1.14}$$

Then it follows that $pv + \frac{a}{v} - \frac{b}{\beta v} = \frac{1}{\beta}$ so that

5.1.15
$$(p + \frac{a}{v^2})\, v = (1 + \frac{b}{v})\frac{1}{\beta} = \frac{1}{1 - \frac{b}{v}}\frac{1}{\beta} + O(\frac{1}{\beta v^2}) \tag{5.1.15}$$

or $(p + \frac{a}{v^2})(v-b)\beta = 1 + O(v^{-2})$, which gives the equation of state up to $O(v^{-2})$ and for $\beta\varepsilon \ll 1$, *i.e.* at high temperature and low density.

It is in fact possible to compute, or at least to give integral representations of the coefficients of arbitrary order of the virial series:

5.1.16
$$\beta p = v^{-1} + \sum_{p=2}^{\infty} c_p(\beta) v^{-p} \tag{5.1.16}$$

and one can even show that the series converges for β small and v large (*i.e.* high temperature and small density) if the stability and temperedness conditions discussed in Chap.II, Chap.IV, (4.1.1) hold.

Equation (5.1.15) can be compared with a well-known empirical equation of state, the Van der Waals equation:

5.1.17
$$\beta(p + a/v^2)(v - b) = 1 \quad \text{or} \quad (p + An^2/V^2)(V - nB) = nRT \tag{5.1.17}$$

where, denoting Avogadro's number by N_A,

5.1.18
$$A = aN_A^2, \quad B = bN_A, \quad R = k_B N_A, \quad n = N/N_A \,. \tag{5.1.18}$$

It is clear that (5.1.16) and (5.1.17) coincide up to quantities of $O(v^{-1})$ hence (once an explicit form for φ like (5.1.1) has been assumed as a good description of the system) (5.1.18), (5.1.14) show us how it is possible to access the microscopic parameters ε and r_0 of the potential φ via measurements detecting deviations from the Boyle-Mariotte law $\beta pv = 1$ of the rarefied gases:

5.1.19
$$\varepsilon = 3a/8b = 3A/8BN_A, \quad r_0 = (3b/2\pi)^{1/3} = (3B/2\pi N_A)^{1/3} \,. \tag{5.1.19}$$

Equation (5.1.17) is, however, empirically used *beyond* its validity region (very large v, *i.e.* very small density) by regarding A, B as phenomenological parameters to be experimentally determined by measuring them near generic values of p, V, T. The result is that the values of A, B do not "usually vary too much" and, apart from this small variability of A, B as functions of v, T, the predictions of (5.1.17) have been in reasonable agreement with experience until, as the precision of the experiments increased over the years, serious inadequacies eventually emerged.

A striking prediction of (5.1.17), taken literally, is that the gas undergoes a "gas-liquid" phase transition with a critical point at a temperature T_c, volume v_c and pressure p_c that can be computed via (5.1.17) and are given by (see §1.2, table (1.1))

5.1.20 $$RT_c = 8A/27B, \qquad V_c = 3B \qquad (n = 1)\,. \tag{5.1.20}$$

The critical temperature is defined as the largest value T_c of the temperature for which the graph of p as a function of v is not monotonic decreasing; the critical volume V_c is the value of v at the horizontal inflection point occurring for $T = T_c$.

At the same time this is very interesting as it shows that there are simple relations among the critical parameters and the microscopic interaction constants ($\varepsilon \simeq k_B T_c$ and $r_0 \simeq (V_c/N_A))^{1/3}$:

5.1.21 $$\varepsilon = 81 k_B T_c/64, \qquad r_0 = (V_c/2\pi N_A)^{1/3} \tag{5.1.21}$$

if the model (5.1.1) is used for the interaction potential φ, see the table in §1.2.

On the other hand, (5.1.17) *cannot be accepted acritically* not only because in its derivation we made various approximations (essentially neglecting $O(v^{-1})$ in the equation of state), but mainly because for $T < T_c$ the function p is *no longer monotonic* in v, and the latter is a thermodynamic function that in Chap.IV and Chap.II has been shown to be monotonic nonincreasing as a consequence of the *very general convexity* of the free energy, evaluated for instance in the canonical ensemble $f_c(\beta, v)$, as a function of v, *i.e.* $\partial^2 f_c/\partial v^2 \geq 0$ so that $-\partial p/\partial v = \partial^2 f_c/\partial v^2 \geq 0$.

If, nevertheless, the isotherms of (5.1.17) are taken seriously even for $T < T_c$, by interpreting them as describing metastable states, then the "correct" equation of state can be obtained by noting that p as a function of v has a horizontal plateau $[v_l, v_g]$ in the situations in the Fig. 5.1.1. Here the plateau associated with the represented isotherm is drawn; hence the density undergoes a jump from v_l to v_g as the pressure decreases and v_l, v_g are interpretable as the specific volumes of the liquid and of the gas.

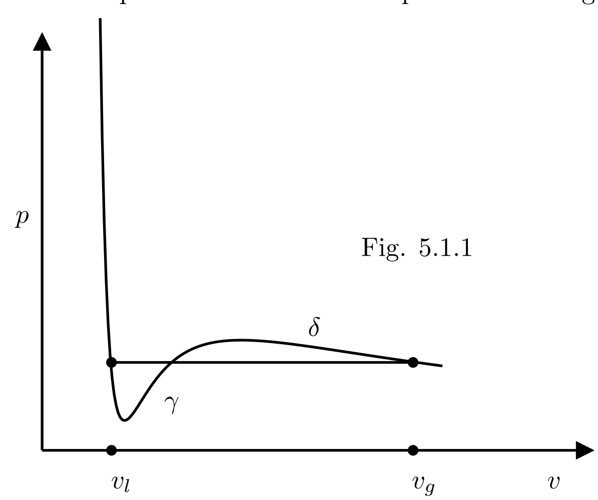


Fig. 5.1.1

The horizontal plateau must be drawn so that the areas γ, δ are equal. The reason is that the reversible thermodynamic *cycle* obtained by having the system go through a sequence of transformations along the plateau and back along the curved parts of the isotherm would yield an output of work represented by the difference between the areas (if run in an appropriate direction). However it would be a Carnot cycle at constant temperature which, by the second principle of Thermodynamics, should instead yield 0 work.

This is the well-known *Maxwell construction* that, as we see, is motivated in a *rather obscure* way because it is not clear whether is is really possible to perform the above Carnot cycle since it is at least doubtful, [LR69], that the *intermediate* states with p increasing with v could be realized experimentally or even be theoretically possible (see, however, the theorem in §5.2).

The van der Waals equation, refined and complemented by Maxwell's rule, nevertheless provides a simple picture for the understanding of the liquid-gas transition in statistical mechanics. *But* it predicts the following behavior:

5.1.22
$$\begin{aligned} &(p-p_c) \propto (V-V_c)^\delta \qquad &&\delta = 3,\, T=T_c \\ &(v_g - v_l) \propto (T_c - T)^\beta \qquad &&\beta = 1/2, \text{ for } T \to T_c^- \end{aligned} \tag{5.1.22}$$

which are in sharp contrast with the experimental data gathered in the twentieth century. For the simplest substances one finds instead $\delta \cong 5$, $\beta \cong 1/3$.

An accurate measurement of δ and β is very delicate and this explains why, for a long time, the equation of van der Waals has been considered a "good representation" not only for a high-temperature low-density gas regime but for the liquid-gas transition regime as well. To gain an idea of the orders of magnitude of the constants A, B, hence of the microscopic interaction data, see the table at the end of §1.2.

One should stress that the disagreement between theory and experiment that we are discussing has a rather different meaning and implications if compared with the discussions in Chap.III. The disagreement here is due to bad approximations (such as having neglected higher-order corrections in v^{-1} in (5.1.15) or such as having assumed that the virial series converged even for values of v, T close to the critical point).

Here the disagreement does not involve fundamental questions on the foundations of the theory: it only involves the analysis of whether a certain approximation is reasonable or correct, or not.

One should remark, last but not least, that the blind faith in the equation of state (5.1.17) is untenable also because of another simple remark: nothing in the above analysis would change if the space dimension was $d = 2$ or $d = 1$: but in the last case, $d = 1$, one can easily *prove* that the system, if the interaction decays rapidly at infinity, does not undergo *phase transitions*, a fact usually known as *Landau's argument*, see §152 in [LL67], and which can be made into a mathematical theorem proved, as such, by van Hove, [VH50], see §5.8 below.

In fact it is now understood that the Van der Waals equation represents *rigorously* only a limiting situation, in which the particles have a hard core interaction (or a strongly repulsive one at close distance) and a further smooth long-range interaction φ: very small but with very long range. This is discussed in §5.2.

As a final comment it is worth stressing that the virial theorem gives in principle the corrections to the equation of state in a rather direct and simple form as time averages of the virial of the internal forces. Since the virial of the internal forces is easy to compute if one knows the positions of the particles as a function of time we see that the theorem *provides a method for computing the equation of state* in numerical simulations. In fact this idea has been exploited in many numerical experiments, in which the (5.1.5) plays a key role.

§5.2. The Modern Interpretation of van der Waals' Approximation

Suppose that the system has an interaction potential $\varphi(r) = \varphi_{hc}(r/r_0) + \gamma^3\varphi_0(\gamma r/r_0)$ where $\varphi_{hc}(r/r_0)$ vanishes for $r > r_0$ and is $+\infty$ for $r \le r_0$ (*hard core potential*), while φ_0 is a smooth function with short range (*i.e.* either eventually equal to 0 as $r \to \infty$, or tending to 0 exponentially fast as $r \to \infty$, say). Here γ is a dimensionless parameter which is really used to set a variable value of the range to $\gamma^{-1}r_0$.

In other words we assume that, apart from the hard core, the particles interact via a potential which is very long range as $\gamma \to 0$ but, at the same time, it becomes very weak.

When γ is very small and the density of the system is fixed to be ρ we see that the energy of interaction between one particle and the remaining ones will be essentially *entirely* due to the particles that are very far apart: the close ones, being (relatively) few, will therefore contribute a small amount to the energy, because the strength of the potential is very weak, proportional to γ^3.

The energy of a particle in the force field of the others is in fact, if $\overline{\varphi}_0$ is the (γ-independent) integral of $\gamma^3\varphi_0(\gamma r/r_0)$

5.2.1 $$\rho\int\gamma^3\varphi_0(\gamma\frac{r}{r_0})\,d^3r \equiv \rho\int\varphi(\frac{r}{r_0})\,d^3r \overset{def}{=} \rho\,\overline{\varphi}_0 \tag{5.2.1}$$

so that the energy of a configuration in which the hard cores of the particles do not overlap will be essentially given, at least for small γ, by

5.2.2 $$U = \frac{1}{2}N\,\rho\,\overline{\varphi}_0\,. \tag{5.2.2}$$

The quantity $\rho\,\overline{\varphi}_0$ is sometimes called the *mean field* at density ρ.

The last relation allows us to compute immediately the canonical partition function. Let $V - Nv_0$ be the volume available to each particle, *i.e.* the

total volume minus the volume occupied by the impenetrable hard cores of the particles: v_0 is of the order r_0^3 and it will be taken, to be in agreement with (5.1.14), to be

5.2.3
$$b = 4v_0 = 4\frac{4\pi}{3}(\frac{r_0}{2})^3. \tag{5.2.3}$$

Then if the energy of a configuration is well approximated by (5.2.2), the canonical partition function is approximately:

5.2.4
$$\begin{aligned} Z(\beta,v) = & \int e^{-\beta(\frac{\underline{p}^2}{2m}+\frac{\rho\,\overline{\varphi}_0}{2}N)}\frac{d^{3N}\underline{q}\,d^{3N}\underline{p}}{h^{3N}N!} = \\ = & \Big(\frac{2\pi m}{\beta h^2}\Big)^{\frac{3N}{2}}\frac{e^{-\beta\frac{\rho\,\overline{\varphi}_0}{2}N}\,(V-Nb)^N}{N!} \end{aligned} \tag{5.2.4}$$

where a somewhat uncontrolled approximation is made about the $\underline{q}$ integrations, as clearly the integral over the configurations of N particles that are constrained to stay at a distance r_0 from each other is a highly non trivial quantity: only naively can one hope to approximate it by $(V-Nb)^N$, even if we allowed simple adjustements of the value of the empirical "excluded volume" b to improve the approximation.

If one accepts (5.2.4) then the equation of state can be computed straightforwardly. In fact the free energy f_c, see (2.3.8), is given by

5.2.5
$$\begin{aligned} -\beta\, f_c(\beta,v) = & \lim_{N\to\infty}\frac{1}{N}\log Z(\beta,v) = \\ = & \log(\sqrt{2\pi m\beta^{-1}h^{-2}})^3 - \beta\frac{\overline{\varphi}_0}{2}v^{-1} + \log(v-b)e \end{aligned} \tag{5.2.5}$$

leading, by differentiation, to

5.2.6
$$\beta\, p(\beta,v) = -(\frac{\partial f_c}{\partial v})_\beta = \frac{\beta\,\overline{\varphi}_0}{2v^2} + \frac{1}{v-b} \tag{5.2.6}$$

which coincides with (5.1.17), thus providing an alternative interpretation of the van der Waals equation and motivating the qualification, which is usually given to it, of *mean field theory*.

The above discussion shows that the van der Waals equation can be exact only if the interaction has extremely long range and at the same time just weak enough to have a nonzero integral $\overline{\varphi}_0$ so that it is correspondingly so small that any individual particle contribution to the potential energy of a fixed particle is negligible apart, of course, from the hard core energy which, unsatisfactorily, is taken into account by replacing the integral over the configurations of non overlapping cores by $(V-Nb)^N$.

In fact the latter approximation can be eliminated by replacing (5.2.4) by

the more accurate:

5.2.7

$$\begin{aligned} Z(\beta,v) &= \int e^{-\beta(\frac{p^2}{2m}+\frac{1}{2}\rho\,\overline{\varphi}_0 N)}\frac{d^{3N}}{h^{3N}N!}\,d\underline{q} = \\ &= \Big(\sqrt{\frac{2\pi m}{\beta h^2}}\Big)^{3N} e^{-\frac{1}{2}\beta\,\rho\,\overline{\varphi}_0\,N} Z_0(v), \\ Z_0(v) &= \int_{|\underline{q}_i-\underline{q}_j|>r_0}\frac{d^{3N}\underline{q}}{N!} \end{aligned} \tag{5.2.7}$$

where the configuration integral $Z_0(v)$ (which is β-independent) over the nonoverlapping hard core configurations $\{\underline{q}_i\}$ in the volume V is not computed.

Equations (5.2.7) then imply that the free energy and the equation of state of our gas are:

5.2.8

$$-\beta f(\beta,v) = -\beta f_0(\,v) - \beta\frac{\overline{\varphi}_0}{2v}, \qquad \big(p(\beta,v) - \frac{\overline{\varphi}_0}{2v^2}\big)\,\beta \,=\, P_0(v) \tag{5.2.8}$$

where

5.2.9

$$P_0(v) \stackrel{def}{=} -\frac{\partial}{\partial v}\frac{1}{N}\log Z_0(v) \tag{5.2.9}$$

is the (temperature-independent) product of β times the pressure p_0 of the hard core gas. It replaces its crude approximation $P_0(v) = \beta p_0(v) = (v-v_0)^{-1}$ in (5.2.6).

It is not difficult to see that the β-independence of $P_0(v)$ implies that, if $\overline{\varphi}_0 < 0$, *i.e.* if the potential has a long range *attractive* (*i.e.* negative) tail, then (5.2.8) will have, at low temperatures, a graph which is qualitatively similar to that of (5.1.17) with $a > 0$ (hence like Fig. 5.1.1).

Thus the equation of state (5.2.8) will show phase transitions, and also the phenomena of *negative compressibility* and *metastability*.

The negative compressibility can be eliminated by Maxwell's rule. But one is still left with the unpleasant feeling that somehow one is doing something wrong. This is clearly signaled by the fact that in spite of the improvements in the approximations we are *still* getting a pressure that is a nonmonotonic function of the specific volume (if β is large enough, *i.e.* if the temperature is low).

At least in one-dimensional gases the excluded volume problem is trivial and one can simply check that $\beta p_0(\,v) = P_0(v)$ is *indeed* $\frac{1}{v-v_0}$ with $v_0 = r_0$, and, therefore, this is clearly a contradiction because we are getting a non monotonic pressure in a situation in which the theory of §4.3 does apply, and implies convexity of the free energy, *i.e.* monotonicity of the pressure.

Continuing to denote by $P_0(v)$ the temperature-independent product $\beta p_0(\beta,v)$ of β times the pressure of the pure hard core gas, the following result sheds a great amount of light on the intricacy of the above situation, showing that the presence of a negative compressibility region is an *artifact* of the mean field approximation:

Theorem: Suppose that we fix $\gamma > 0$ and we call $p(\beta, v; \gamma)$ the canonical pressure (in the thermodynamic limit) for the gas interacting with the potential

5.2.10
$$\varphi(r) = \varphi_{hc}\big(\frac{r}{r_0}\big) + \gamma^3 \varphi_0\big(\gamma \frac{r}{r_0}\big) \tag{5.2.10}$$

where $\varphi_{hc} = 0$ for $r > r_0$, $\varphi_{hc} = \infty$ for $r \le r_0$; φ_0 is a smooth potential rapidly decreasing at ∞ and with integral $\overline{\varphi}_0 < 0$. Let also $P_0(v)$ be the (β-independent) product of β times the pressure of the pure hard core gas (in the thermodynamic limit), i.e. the pressure of the gas in which the particles interact only via the hard core potential φ_0. Then:

5.2.11
$$\beta\, p(\beta, v) \stackrel{def}{=} \lim_{\gamma \to 0} \beta\, p(\beta, v; \gamma) = \Big[\frac{\beta\, \overline{\varphi}_0}{2v^2} + P_0(v)\Big]_{Maxwell\ rule} \tag{5.2.11}$$

where the subscript "Maxwell rule" means that in the regions where the right-hand side is not monotonic in v (existent if $\overline{\varphi}_0 < 0$ and β is large) the pressure $p(\beta, v)$ is obtained with the help of the Maxwell construction discussed in §5.1.

Thus we see that *there is indeed a firm foundation to Maxwell's rule* which does not rest on *dubious* Carnot cycles: the van der Waals equation becomes rigorously valid in the limit in which the attractive tail of the potential becomes very weak but with so long-range that the *mean potential* ("mean field"), see (5.2.2), that it generates in a point has a fixed value.

If the dimension is $d = 1$ the hard core gas pressure $P_0(v)$ is rigorously $\beta p_0 = (v - v_0)^{-1}$ and the equation of state becomes *exactly* the Van der Waals equation. In higher dimensions $\beta p_0 = (v - v_0)^{-1}$ is only an approximation (no matter how v_0 is chosen), but the basic fact that the equation of state is a trivial modification of a reference, "simpler" (so to speak), system (the hard core gas) together with Maxwell's rule remains valid.

One can also say that the van der Waals equation arises when one interchanges two limits: the thermodynamic limit and the limit of infinite range $\gamma \to 0$. It is obvious that if instead of taking the limit $V \to \infty$ first and *then* the limit $\gamma \to 0$ one did the opposite then the equation of state would have been $p = p_0$ and the attractive tail would have given *no contribution.*

The potentials like (5.2.10) are called *Kac's potentials*, [HKU63] and one can say that the above theorem plays a role analogous to that of Lanford's theorem for the Boltzmann equation, see §1.8 and §1.9: in both cases a statement that has approximate validity becomes exact in a suitable limit. And in both cases the statement seems incompatible with obvious properties of the system (reversibility in the first case and strict convexity in finite volume systems of the free energy in the second), although of course the first case concerns a far more fundamental problem than the second.

But both cases are instances of a method of analysis that has been developed very much in the twentieth century, in which one tries to understand some properties that cannot be exactly true in normal situations but that become exactly true in suitable limiting situations thus leading to a more

or less detailed understanding of why they *may look* true even when the limit is not taken. Of course a complete theory should also come together with estimates (and possibly reasonable ones) of how far we are, in concrete situations, from the limiting cases (*i.e.* how big are the corrections on the quantities on which we might be interested).

The method is a modern interpretation of the basic conception of Boltzmann on the relation between the apparent continuum of reality, as we perceive it and input it in most of our models or theories for its interpretation and understanding, and the possibly intrinsic and deep discrete nature of reality and of our own thinking.

This is exemplified by the quotation in §1.1 and in many others among Boltzmann's writings, for instance:

"*The concepts of differential and integral calculus separated from any atomistic idea are truly metaphysical, if by this we mean, following an appropriate definition of Mach, that we have forgotten how we acquired them*", p. 56 in [Bo74].

And I cannot resist the temptation of more quotations, as this is really music for the mind:

"*Through the symbols manipulations of integral calculus, which have become common practice, one can temporarily forget the need to start from a finite number of elements, that is at the basis of the creation of the concept, but one cannot avoid it*"; p. 55 in [Bo74].

or:

"*Differential equations require, just as atomism does, an initial idea of a large finite number of numerical values and points ...... Only afterwards it is maintained that the picture never represents phenomena exactly but merely approximates them more and more the greater the number of these points and the smaller the distance between them. Yet here again it seems to me that so far we cannot exclude the possibility that for a certain very large number of points the picture will best represent phenomena and that for greater numbers it will become again less accurate, so that atoms do exist in large but finite number*, see p. 227 in [Bo74];

and:

"*This naturally does not exclude that, after we got used once and for all to the abstraction of the volume elements and of the other symbols* [of calculus] *and once one has studied the way to operate with them, it could look handy and luring, in deriving certain formulae that Volkmann calls formulae for the coarse phenomena, to forget completely the atomistic significance of such abstractions. They provide a general model for all cases in which one can think to deal with* 10^{10} *or* $10^{10^{10}}$ *elements in a cubic millimeter or even with billions of times more; hence they are particularly invaluable in the frame*

of Geometry, which must equally well adapt to deal with the most diverse physical cases in which the number of the elements can be widely different. Often in the use of all such models, created in this way, it is necessary to put aside the basic concept, from which they have overgrown, and perhaps to forget it entirely, at least temporarily. But I think that it would be a mistake to think that one could become free of it entirely."[2]

The latter sentence, p. 55 in [Bo74], reminds us that the evaluation of the corrections is of course a harder problem, which it would be a mistake to set aside, even in the above case of mean field theory. In fact the corrections are quite important and somehow even more important than the mean field theory itself, which will remain as a poor idealization of far more interesting cooperative phenomena.

It should also be noted that the above analysis does not allow us to solve a fundamental question: can classical statistical mechanics predict and describe phase transitions? We have seen that the van der Waals theory is no proof that *when no infinite range mean field limit* is taken (*i.e.* in most interesting cases) then a system can show phase transitions.

This clearly emerges from the remark that if γ is *not* 0 then in $d = 1$ one can prove, as a theorem (see above, and §5.8), that the system cannot undergo any phase transition whatsoever (and the pressure is strictly monotonic in the specific volume, "no plateau" at all); and nevertheless the above theorem also holds in $d = 1$, where in fact it leads precisely to the Van der Waals equation (with the volume of the box being replaced by the length of the box and similar obvious changes).

It is therefore important to see whether genuinely short-range models (no γ around) generate equations of state with phase transitions. This will be thoroughly discussed in Chap.VI in simplified models, because in the cases in which one would like to have results the problem is *still open*, and we shall see that in the simplified models phase transitions are possible even when the interactions have short range and the analysis will leave little doubt (in fact no doubt at all) that phase transitions are possible in classical statistical mechanics, without the necessity of introducing any new assumptions or new physical laws.

We should however mention that important breakthroughs seem to be under way: see [Jo95] and [LMP98].

The standard approach to the van der Waals theory (also called mean field theory) can be found in [CC53], p. 284. A more refined and interesting formulation is in [VK64]. A precise and very clear theory is in [LP66]. The first precise understanding (and full proof in particular cases) of mean field theory is in [HKU63], in a series of papers reproduced, with introductory remarks, in [LM66]. A more phenomenological but very interesting and original theory is in the book [Br65], where the most common phase transitions are treated from the unifying point of view of the mean field theory. The original work of van der Waals has been reprinted, [VW88].

---

[2] Lucretius would not have said it better.

§5.3. Why a Thermodynamic Formalism?

In the next sections we devote ourselves to a more detailed analysis of the framework in which phase transitions could be placed and of the techniques that one may envisage to apply towards an understanding of their properties. It will be a rather abstract analysis, usually called "*thermodynamic formalism*" that plays in statistical mechanics a role akin to that played by the Hamiltonian formalism in mechanics.

One does not have to recall that the formalism of Hamilton, in itself, does not make mechanics problems any easier than other formalisms. However it has become a tautology that it is a very appropriate formalism to describe mechanical phenomena. The same can, or should, be said about the thermodynamic formalism.

The theory of orthodic ensembles provides us with a model of Thermodynamics but, strictly speaking, only in the limit of infinite volume. In this situation one also obtains equivalence between the various ensembles, see Chap.II.

The elements of the orthodic ensembles describe in great detail the structure of the *thermodynamic phases* (*i.e.* macroscopic states), well beyond the simple microscopic definition of the classical thermodynamic quantities, and even provide us with the (surprising) possibility of computing theoretically some relations between them (*e.g.* the equation of state). Every element of a statistical ensemble describes details of the microscopic configurations that are typical of the corresponding phase, because it gives the probability of each individual microscopic configuration.

The problem of the "thermodynamic limit" theory is that of establishing a formalism in which it becomes possible to make precise and sharp various statements that we have made so far, on intuitive or heuristic grounds, and thus lay the grounds for a deeper analysis and for deeper physical questions.

We shall only consider the case of classical statistical mechanics, in which one neglects the size of Planck's constant h.

What follows, as stated at the end of the previous section, is a *formalism*: as with all formalisms it has interest only because it provides a natural frame (as experience taught us) in which the discussion of the most important questions and applications can be situated. This is not the place to argue that this is the best formalism: others are possible and in the end equivalent. But we need a formalism just in order to formulate precise questions, suitable of being given quantitative answers. The amount of work to be done will be independent of the formalism used (of course).

It is well known that for each class of problems the formalism in which they are formulated often has a clarifying and unifying role: the emergence of a "good" formalism is often successive to the solution of important problems in the field. This seems to be the case of the thermodynamic formalism and the following few sections should be understood from this viewpoint.

As an example of the problems that it would be premature to formulate without a clear formalism in which they fit one can quote:

(1) describing the spatial correlations between particles in a gas,

(2) describing (and in fact defining) the surfaces of separation between different, but coexisting, phases,

(3) understanding the formation and dissociation of gas molecules or atoms into their more primitive constituents in stationary state situations, and other cooperative phenomena.

A initial question is in which sense an element of a statistical ensemble describes a probability distribution on phase space once the limit of infinite volume has been taken. We shall consider here only the grand canonical ensemble representation of the equilibrium states, because it is somewhat easier to discuss than the canonical or microcanonical. We examine the case of a system of identical particles with mass m enclosed in a cubic box V.

The particles interaction will be assumed to take place via a potential φ satisfying at least the stability and temperedness conditions (4.1.1), that are necessary according to the analysis of Chap.IV in the theory of the ensembles: *i.e.* $\Phi(\underline{q}_1,\ldots,\underline{q}_n)=\sum_{i<j}\varphi(\underline{q}_i-\underline{q}_j)\ge -Bn$ ("stability") and $|\varphi(\underline{r})|<C|\underline{r}|^{-(3+\varepsilon)}$ for $|\underline{r}|\ge r_0>0$ ("temperedness"), $B,C,\varepsilon>0$.

To avoid several technical problems it will also be convenient to suppose that the potential φ has a *hard core* with diameter r_0, *i.e.* it is *defined* as the sum of a smooth potential plus a singular potential which is $+\infty$ for $|\underline{r}|<r_0$ so that $\varphi(\underline{r})=+\infty$ for $|\underline{r}|<r_0$. This has the physical significance that two particles cannot be closer than r_0, but a large part of what will be discussed does apply, with suitable modifications (and several open problems left), to the case of a *superstable potential*. This is a potential such that there are two constants $A,B>0$ such that

5.3.1 $$\Phi(\underline{q}_1,\ldots,\underline{q}_n)\ge -Bn+An^2/V\quad\text{if } \underline{q}_1,\ldots,\underline{q}_n\in V \tag{5.3.1}$$

where V is an arbitrary cubic volume containing an arbitrary number $n\ge 2$ of particles located at $\underline{q}_1,\ldots,\underline{q}_n$. The Lennard-Jones potential, see (5.1.1), is a typical example of a superstable potential. However the potential $\varphi=0$, the free gas model, is *not* superstable (although it is trivially stable), see [Ru70] for a general theory of such potentials.

Let V be a cubic volume and consider the element $\mu^{(\beta,\lambda,V)}$ of the grand canonical ensemble with parameters (β,λ) and with particles confined in V: $\beta=1/k_BT$, $k_B=$ Boltzmann's constant and $T=$ temperature, $\lambda=$ chemical potential, see §2.5. The probability of finding n particles in the microscopic state $d\underline{p}_1\ldots d\underline{p}_n\ d\underline{q}_1\ldots d\underline{q}_n$ in the distribution $\mu^{(\beta,\lambda,V)}$ is:

5.3.2 $$\mu^{(\beta,\lambda,V)}(d\underline{p}_1\ldots d\underline{q}_n)=\frac{e^{-\beta(E(\underline{p},\underline{q})+\lambda n)}}{\Xi(\beta,\lambda)}\,\frac{d\underline{p}_1\ldots d\underline{p}_n\,d\underline{q}_1\ldots d\underline{q}_n}{n!\,h^{3N}} \tag{5.3.2}$$

where $E(\underline{p},\underline{q})=T(\underline{p})+\Phi(\underline{q})=\sum_{i=1}^n\underline{p}_i^2/2m+\Phi(\underline{q}_1,\ldots,\underline{q}_n)$ and Ξ is the grand canonical partition, see §2.5. We want to take the limit of (5.3.2) as

$V \to \infty$ and interpret it as a probability distribution on the infinite system configurations that one reaches in this way.

One begins by giving a precise definition of the infinite system configurations; then comes the problem of giving a meaning to the limit as $V \to \infty$ of (5.3.2) and finally one will want to characterize the distributions that are found by following this limiting procedure starting from (5.3.2), or starting from the more general grand canonical distributions with fixed external particle boundary conditions; the latter were introduced, as generalization of (5.3.2), in §2.5: see (2.5.2).

§5.4. Phase Space in Infinite Volume and Probability Distributions on it. Gibbs Distributions

It is natural to define the phase space $\overline{M}$ in infinite volume as the space of the *sequences* $(\underline{p}, \underline{q}) = (\underline{p}_i, \underline{q}_i)_{i=1}^{\infty}$ of momenta and positions such that in every finite (cubic) volume there are only finitely many particles, called *locally finite configurations*: if we consider systems of particles with hard core of diameter $r_0 > 0$ this will be "automatic", as the only configurations $\underline{q}$ that we have to consider are those with $|\underline{q}_i - \underline{q}_j| > r_0$, for $i \neq j$.

However, to take into account microscopic indistinguishability the configuration space will not be $\overline{M}$, but the space M obtained from $\overline{M}$ by identifying sequences $(\underline{p}, \underline{q})$ differing by a permutation of the particles.

A *probability distribution* on M is in general defined so that the following question makes sense: what is the probability that by looking in a given (cubic) volume V^0 one finds in it exactly g particles with momenta in $d\underline{p}_1 \ldots d\underline{p}_g$ and positions in $d\underline{q}_1 \ldots d\underline{q}_g$?

Therefore the probability distribution μ will be characterized by the functions $f_{V^0}(\underline{p}_1, \ldots, \underline{p}_g, \underline{q}_1, \ldots, \underline{q}_g)$ such that the quantity:

5.4.1
$$f_{V^0}(\underline{p}_1, \ldots, \underline{p}_g, \underline{q}_1, \ldots, \underline{q}_g) \frac{d\underline{p}_1 \ldots d\underline{p}_g \, d\underline{q}_1 \ldots d\underline{q}_g}{g!} \tag{5.4.1}$$

is the probability just described. The functions f_{V^0} will be called the *local distributions* of μ: the factor $g!$ could be included in f_{V^0}, but it is customary not to do so since the particles are indistinguishable and this factor simplifies combinatorial considerations.

By using the functions f_{V^0} it will be possible to evaluate the average value of a *localized observable*, localized inside the volume V^0: this is, by definition, a function on phase space that depends on $(\underline{p}, \underline{q}) \in M$ *only* via the state of the particles located in V^0. Adopting the convention that $\{\underline{p}\}_g, \{\underline{q}\}_g \stackrel{def}{=} \underline{p}_1, \ldots, \underline{p}_g, \underline{q}_1, \ldots, \underline{q}_g$, if F is such a local observable we can write its average as

5.4.2
$$\langle F \rangle = \sum_{g=0}^{\infty} \int_{R^{3g} \times V_0^g} F(\{\underline{p}\}_g, \{\underline{q}\}_g) \cdot f_{V^0}(\{\underline{p}\}_g, \{\underline{q}\}_g) \frac{\mathrm{d}\{\underline{p}\}_g \, \mathrm{d}\{\underline{q}\}_g}{g!} \,. \tag{5.4.2}$$

Consider a probability distribution like (5.3.2) describing a particle system enclosed in a "large global" volume V that we suppose cubic. With a fixed $V^0 \subset V$ (we think here of V as huge and V^0 as much smaller), one can compute the probability that inside V^0 the configuration $(\underline{p}, \underline{q})$ will consist in g particles in $d\underline{p}_1 \ldots d\underline{p}_g d\underline{q}_1 \ldots d\underline{q}_g$. Once the appropriate integrals are performed one will find necessarily an expression like

5.4.3
$$f^{(V)}_{V^0}(\underline{p}_1, \ldots, \underline{p}_g\,, \underline{q}_1, \ldots, \underline{q}_g)\, \frac{d\underline{p}_1 \ldots d\underline{p}_g\, d\underline{q}_1 \ldots d\underline{q}_g}{g!}\,. \tag{5.4.3}$$

It is then natural to define the limit as $V \to \infty$ of the probability distribution $\mu^{(\beta,\lambda,V)}$, (5.3.2),[3] as the distribution μ on M characterized by the *local distributions*

5.4.4
$$f_{V^0}(\underline{p}_1, \ldots, \underline{p}_g, \underline{q}_1, \ldots, \underline{q}_g) = \lim_{V \to \infty} f^{(V)}_{V^0}(\underline{p}_1, \ldots, \underline{p}_g, \underline{q}_1, \ldots, \underline{q}_g) \tag{5.4.4}$$

provided the limit exists for each V^0.

It can be shown that if the interparticle potential φ is superstable, see (5.3.1), and hence *a fortiori* if it has a hard core, then the limit (5.4.4) exists, at least along subsequences of any sequence of volumes V with $V \to \infty$. In the hard core case this is an almost obvious "compactness argument" (*i.e.* a "free" argument based on abstract nonsense).[4]

The same remains true if $\mu^{(\beta,\lambda,V)}$ is replaced by a more general element of a grand canonical ensemble with fixed external particle boundary conditions, provided the external particle density "does not grow too fast with their distance to the origin", see §2.5.

The latter condition means that, fixing a length unit ℓ (arbitrarily), the number $n(\Delta)$ of external particles in a box Δ with side size ℓ does not grow too fast with the distance $d(\Delta, O)$ of Δ from the origin, *e.g.* it satisfies

5.4.5
$$\frac{n(\Delta)}{d(\Delta, O)} \xrightarrow[d(\Delta,O)\to\infty]{} 0. \tag{5.4.5}$$

This condition is automatically satisfied if the interaction has hard core; if it is not satisfied then it is not difficult to find a configuration of external particles such that the above limit does not exist (or is "unreasonable").[5]

---

[3] When one imagines the volume of the "global" container increasing to ∞, keeping the size of the region V^0 that is under scrutiny fixed.

[4] On the contrary it is highly nontrivial, when true, to prove the existence of the limit $V \to \infty$ *without* restricting V to vary along a "suitable" subsequence.

[5] Consider in fact a system of particles interacting via a Lennard-Jones potential, (5.1.1), which is $< -b$ for distances between a and $2a$. Let V be a cubic container and distribute outside V, at distance exactly a, $M = N^c$ external particles, $c > 2/3$. Inside the volume V suppose that there are N particles with $N = \rho V$, $\rho > 0$. With such boundary condition the canonical distribution has a thermodynamic limit in which there are *no particles* in any finite region with probability 1 (*i.e.* $f_{V^0} \equiv 0$ for all V^0). In fact one checks that, putting all N internal particles in a corridor of width a around the boundary, one gets a set of configurations with energy $\le O(-NN^cV^{-2/3})$ and phase space volume $O((aV^{2/3})^N N!^{-1})$. Then a comparison argument similar to those of §4.1 to study the various catastrophes applies. Likewise one can discuss the corresponding example in the grand canonical ensemble (or the microcanonical).

Therefore we can define the *set of Gibbs distributions* on the phase space M as the set of all possible distributions that are obtainable as limits of convergent subsequences, in the sense of (5.4.4), of grand canonical distributions $\mu^{(\beta,\lambda,V)}$ with periodic boundary conditions or with fixed external particle boundary conditions whose density does not grow too fast at infinity in the above sense.

The distributions that are obtained in this way will define the *equilibrium phases* of the system (see §2.5) and are *not* necessarily invariant under translations, *i.e.* such that for every displacement $\xi \in R^3$:

5.4.6
$$f_{V^0+\xi}(\underline{p}_1,\ldots,\underline{p}_g,\underline{q}_1+\xi,\ldots,\underline{q}_g+\xi) = f_{V^0}(\underline{p}_1,\ldots,\underline{p}_g,\underline{q}_1,\ldots,\underline{q}_g)\,. \tag{5.4.6}$$

Except in the (*important*) special case in which periodic boundary coonditions are used, translation invariance symmetry is *broken* by the fact that the system is, before the thermodynamic limit $V \to \infty$, enclosed in a finite box V; and it is not necessarily true that the invariance is “restored” by the mere fact that we send $V \to \infty$.

The physical phenomenon related to the above (possible) *spontaneous breakdown* of translation symmetry is the possibility of the existence of thermodynamic states in which *pure* phases coexist occupying, for instance, each half of the total space allotted to the system, being separated by a microscopically well-defined surface: one should think here of a liquid in equilibrium with its vapor.

Therefore we shall distinguish the set of Gibbs distributions $G^0(\beta,\lambda)$ from its subset $G(\beta,\lambda) \subset G^0(\beta,\lambda)$ consisting in the distributions which are invariant under translations, *i.e.* which have local distributions satisfying (5.4.6).

If μ is a translation-invariant probability distribution on M and if $S = (S_1, S_2, S_3)$ are the translations by one length unit, in the three directions (*i.e.* $S_\alpha(\underline{p}_i,\underline{q}_i)_{i=1}^\infty = (\underline{p}_i,\underline{q}_i+\underline{e}^\alpha)$ with $\underline{e}^\alpha$, $\alpha = 1,2,3$ being the unit vector in the α-th direction), then the triple (M, S, μ) is, according to a well established terminology, a *dynamical system*; this is a useful fact to bear in mind, as we shall see on several occasions.

One could, of course, define the Gibbs distributions by starting from distributions of the canonical ensemble (or microcanonical or any other orthodic ensemble) with fixed external particle boundary conditions.

By so doing one would generate the problem of the equivalence of the ensembles (see Chap.II) in the sense that one should show that the totality of the Gibbs distributions on M built starting from the grand canonical ensemble distributions with fixed external particles boundary conditions does coincide with the totality of the Gibbs distributions (on M) built starting with microcanonical or canonical distributions with periodic or fixed external particle boundary conditions.

The analysis of the latter question is difficult: it is essentially complete only in the case of hard core systems, [Do68a],[Do72], [LR69]; but it is somewhat incomplete in the “general” case of superstable potentials, see [Ru70],[La72], [Ge93]. Nevertheless there is no evidence that there might be conceptual problems on such matters.

§5.5. Variational Characterization of Translation Invariant Gibbs Distributions

If we restrict our attention to the translation-invariant Gibbs distributions $\mu \in G(\beta, \lambda)$, also called *homogeneous phases*, then an alternative and interesting variational characterization of them is often possible.

The first simple remark, that stems immediately from (5.4.4), or from its variants with different fixed external particles boundary conditions, is that if $\mu \in G^0(\beta, \lambda)$ then the momenta distribution is *Maxwellian*, *i.e.* $f_{V^0}(\underline{p}_1, \ldots, \underline{p}_g, \underline{q}_1, \ldots, \underline{q}_g)$ can be written as

5.5.1
$$f_{V^0}(\underline{p}_1, \ldots, \underline{p}_g, \underline{q}_1, \ldots, \underline{q}_g) = \frac{e^{-\beta \sum_{i=1}^{g} \underline{p}_i^2/2m}}{\sqrt{2\pi m \beta^{-1}}^{3g}} \overline{f}_{V^0}(\underline{q}_1, \ldots, \underline{q}_g) \qquad (5.5.1)$$

where the factor in the square root is introduced because it provides an obviously convenient normalization, making $\overline{f}$ a quantity with the dimension of an inverse length to the power $3g$ since the quantity

5.5.2
$$\gamma(\beta) = \sqrt{2\pi m \beta^{-1}} \qquad (5.5.2)$$

is a "momentum". Sometimes one defines instead $\gamma(\beta) = \sqrt{2\pi m \beta^{-1} h^{-2}}$ including in it also the factor h^{-3g} that appears in (2.2.1); with this choice $\overline{f}_{V_0}$ would be dimensionless.

The probability distributions on phase space M with local distributions that depend upon the momenta as in (5.5.1) are called *Maxwellian distributions*. The problem is therefore that of characterizing $\overline{f}_{V^0}$ so that the distribution defined by (5.5.1) is in $G(\beta, \lambda)$.

Going back to a finite total volume a well-known argument shows that (5.3.2) satisfy a *variational principle*. More precisely let $(\underline{p}, \underline{q})$ abbreviate $(\underline{p}_1, \ldots, \underline{p}_n, \underline{q}_1, \ldots, \underline{q}_n)$, and write (5.3.2) as

5.5.3
$$f(\underline{p}_1, \ldots, \underline{p}_n, \underline{q}_1, \ldots, \underline{q}_n)\, \mathrm{d}\underline{p}_1 \ldots \mathrm{d}\underline{p}_n\, \mathrm{d}\underline{q}_1 \ldots \mathrm{d}\underline{q}_n \equiv f_n(\underline{p}, \underline{q})\, \mathrm{d}\underline{p}\, \mathrm{d}\underline{q} \qquad (5.5.3)$$

and set

5.5.4
$$E_n(\underline{p}, \underline{q}) \equiv \sum_{i=1}^{n} \frac{\underline{p}_i^2}{2m} - \Phi(\underline{q}_1, \ldots, \underline{q}_n) \equiv T_n(\underline{p}) + \Phi_n(\underline{q})\,. \qquad (5.5.4)$$

Then consider the functional $J(f) = S(f) - \beta U(f) - \beta\lambda N(f)$ defined on the functions f:

5.5.5
$$\begin{aligned} J(f) \stackrel{def}{=} & \frac{1}{V} \sum_{n=0}^{\infty} \int -f_n(\underline{p}, \underline{q}) \log f_n(\underline{p}, \underline{q}) \frac{\mathrm{d}\underline{p}\, \mathrm{d}\underline{q}}{n!} - \\ & - \frac{\beta}{V} \sum_{n=0}^{\infty} \int f_n(\underline{p}, \underline{q})(E_n(\underline{p}, \underline{q}) + \lambda n) \frac{\mathrm{d}\underline{p}\, \mathrm{d}\underline{q}}{n!} = \qquad (5.5.5) \\ & = - \frac{1}{V} \sum_{n=0}^{\infty} \int f_n(\underline{p}, \underline{q})(\log f_n(\underline{p}, \underline{q}) + \beta E_n(\underline{p}, \underline{q}) + \beta\lambda n) \frac{\mathrm{d}\underline{p}\, \mathrm{d}\underline{q}}{n!}\,. \end{aligned}$$

By the Lagrange multiplier method one checks that $J(f)$ is stationary (actually a maximum) on the set of the $f \geq 0$ such that

5.5.6
$$\sum_{n=0}^{\infty} f_n(\underline{p}, \underline{q}) \frac{\mathrm{d}\underline{p}\, \mathrm{d}\underline{q}}{n!} = 1 \tag{5.5.6}$$

if, still with the notation $(\underline{p}, \underline{q}) = (\underline{p}_1, \ldots, \underline{p}_n, \underline{q}_1, \ldots, \underline{q}_n)$, the functions f_n satisfy:

5.5.7
$$(-\log f_n(\underline{p}, \underline{q}) - \beta(E_n(\underline{p}, \underline{q}) + \lambda n)) = \text{ constant} \tag{5.5.7}$$

i.e. if f is given by (5.3.2).

It is natural, at this point, to introduce the space of all translation invariant distributions μ on M that have a Maxwellian momentum distribution and to define on this space the following functionals: the "specific volume", the "total energy" and the "potential energy" corresponding to the interparticle potential φ, and the "entropy".

We denote such functionals by $v(\mu), u_\varphi(\mu)$, $\overline{u}_\varphi(\mu)$ and $s(\mu)$, respectively, and we write them first in the general case and then we shall consider the expression that they assume when the f_{V^0} have the Maxwellian form (5.5.1).

To simplify we also abbreviate the notation for the local distributions in the volume V^0, see (5.5.1), as:

5.5.8
$$\begin{aligned} f_{V^0,g}(\underline{p}, \underline{q}) &\equiv f_{V^0}(\underline{p}_1, \ldots, \underline{p}_g, \underline{q}_1, \ldots, \underline{q}_g) \\ \overline{f}_{V^0,g}(\underline{q}) &= \overline{f}_{V^0}(\underline{q}_1, \ldots, \underline{q}_g) \end{aligned} \tag{5.5.8}$$

where $(\underline{p}, \underline{q})$ stands for $(\underline{p}_1, \ldots, \underline{p}_g, \underline{q}_1, \ldots, \underline{q}_g)$ and $\mathrm{d}\underline{p}\, \mathrm{d}\underline{q} = \mathrm{d}\underline{p}_1 \ldots \mathrm{d}\underline{p}_g$ $\mathrm{d}\underline{q}_1 \ldots \mathrm{d}\underline{q}_g$. Then the specific volume of μ will be defined by

5.5.9
$$\begin{aligned} v(\mu)^{-1} &= \lim_{V^0 \to \infty} \frac{1}{V^0} \sum_{g=0}^{\infty} \int g f_{V^0,g}(\underline{p}, \underline{q}) \frac{\mathrm{d}\underline{p}\, \mathrm{d}\underline{q}}{g!} = \\ &= \lim_{V^0 \to \infty} \frac{1}{V^0} \sum_{g=0}^{\infty} \int g \overline{f}_{V^0,g}(\underline{q}) \frac{\mathrm{d}\underline{q}}{g!}\,. \end{aligned} \tag{5.5.9}$$

With the notations in (5.5.4) the total energy will be

5.5.10
$$\begin{aligned} u_\varphi(\mu) = &\lim_{V^0 \to \infty} \frac{1}{V^0} \sum_{g=0}^{\infty} \int (T_g(\underline{p}) + \Phi_g(\underline{q}))\, f_{V^0,g}(\underline{p}, \underline{q}) \frac{\mathrm{d}\underline{p}\, \mathrm{d}\underline{q}}{g!} = \\ &= \lim_{V^0 \to \infty} \frac{1}{V^0} \sum_{g=0}^{\infty} \int (\frac{3}{2\beta} g + \Phi_g(\underline{q}))\, \overline{f}_{V^0,g}(\underline{q}) \frac{\mathrm{d}\underline{q}}{g!} \equiv \\ &\equiv \frac{3}{2\beta} v(\mu)^{-1} + \overline{u}_\varphi(\mu) \end{aligned} \tag{5.5.10}$$

where in the first step the gaussian, hence trivial, integrals over the momenta $\underline{p}$ have been performed explicitly and in the second step we used (5.5.9).

Likewise the thermodynamic entropy is:

$$\begin{aligned} s(\mu) &= \lim_{V^0\to\infty} -\frac{1}{V^0}\sum_{g=0}^{\infty}\int f_{V^0,g}(\underline{p},\underline{q})\,\log f_{V^0,g}(\underline{p},\underline{q})\frac{\mathrm{d}\underline{p}\,\mathrm{d}\underline{q}}{g!} = \\ &= \lim_{V^0\to\infty} -\frac{1}{V^0}\sum_{g=0}^{\infty}\int \frac{e^{-\beta T_g(\underline{p})}}{\sqrt{2\pi m\beta^{-1}h^{-2}}^{3g}}\,\overline{f}_{V^0,g}(\underline{q})\cdot \\ &\quad\cdot(-\beta T_g(\underline{p}) - \frac{3g}{2}\log(2\pi m\beta^{-1}h^{-2}) + \log\,\overline{f}_{V^0,g}(\underline{q}))\frac{d\underline{q}\,d\underline{p}}{g!} = \\ &= \lim_{V^0\to\infty} -\frac{1}{V^0}\sum_{g=0}^{\infty}\int \overline{f}_{V^0,g}(\underline{q})\frac{\mathrm{d}\underline{q}}{g!\gamma(\beta)^{3g}}\cdot \\ &\quad\cdot(-\frac{3g}{2} - \frac{3g}{2}\log(2\pi m\beta^{-1}h^{-2}) + \log\overline{f}_{V^0,g}(\underline{q})) \equiv \\ &\equiv -\,v(\mu)^{-1}\,\log(2\pi\,e\,m\beta^{-1}h^{-2})^{3/2} + \overline{s}(\mu)\,. \end{aligned}$$

5.5.11 (5.5.11)

All limits above do exist in the case of systems with hard core potentials: to prove this the techniques are similar to those used in Chap.IV to discuss the existence of the thermodynamic limit. The limits, however, exist under much more general conditions that we shall not discuss here.

We now maximize, on the space of the Maxwellian translation-invariant distributions μ on M (with inverse temperature β, *i.e.* having the form (5.5.1)), the functional:

5.5.12
$$s(\mu) - \beta\lambda v(\mu)^{-1} - \beta u_\varphi(\mu) \qquad (5.5.12)$$

and let $\beta\, p(\beta,\lambda)$ denote the supremum of (5.5.12).

We proceed by *quoting only results that are valid in the case of hard core systems, to avoid discussions on the more general superstable case* (for which similar, but less satisfactory results can be obtained), the general discussion being somewhat technical, [Ru70]. If φ has a hard core one has:

5.5.13
$$\beta\, p(\beta,\lambda) = \max_{\mu}\big(s(\mu) - \beta\lambda v(\mu)^{-1} - \beta u_\varphi(\mu)\big) \qquad (5.5.13)$$

and the maximum is reached exactly on the translation-invariant Gibbs distributions $\mu\in G(\beta,\lambda)$, and only on them; see [Do68a],[LR69].[6] One can check that the meaning of the maximum value $p(\beta,\lambda)$ is that of "pressure" (leaving aside mathematical rigor this is, in fact, quite clear from the discussions in §2.5 and above).

The variational property (5.5.13) has been heuristically based on the mentioned check (see (5.5.5),(5.5.7)) that the functional (5.5.5) leads, in a finite

---

[6] In the quoted papers one considers lattice systems, see §5.10 below, but the techniques and results can be extended to hard-core systems quite straightforwardly.

volume container with "open boundary conditions" (*i.e.* no fixed particles outside the container), to the element (β, λ) of the grand canonical ensemble. However a remarkable feature of (5.5.13) is that *nevertheless* it happens that the solutions of the maximum problem (5.5.12) contain, as well, the translation invariant Gibbs distributions that can be obtained by imposing *general* fixed external particle boundary conditions and subsequently considering the thermodynamic limit of the distributions so obtained, [Do68a], [Do68b], [Do72], [LR69].

It can be checked (this is a simple but nontrivial theorem) that the set of translation-invariant probability distributions that realize the maximum in (5.5.13) (*i.e.* the set denoted $G(\beta, \lambda)$ of the Gibbs distributions with inverse temperature β and chemical potential λ) form a *convex set* (*i.e.* $\mu_1, \mu_2 \in G(\beta, \lambda)$ implies $a\mu_1 + (1-a)\mu_2 \in G(\beta, \lambda)$ for all $a \in (0, 1)$). Furthermore the convex set is actually a *simplex*, *i.e.* such that every $\mu \in G(\beta, \lambda)$ can be represented *uniquely* as a convex superposition of *extremal distributions* in $G(\beta, \lambda)$.

The statistical interpretation of a (convex) superposition of two probability distributions is that of a *mixture*: hence the meaning of the latter described property is interesting. It says, in other words, that if the extremal distributions of $G(\beta, \lambda)$ are interpreted as the *pure homogeneous* (*i.e.* translation invariant) *phases*, then all the other elements ("homogeneous phases") in $G(\beta, \lambda)$ are *mixtures* of pure phases and they can be represented as such in a unique way.

For instance if $G(\beta, \lambda)$ contains only two extremal elements μ_+ and μ_-, the first representing the "liquid phase" and the second the "gaseous phase", then every other distribution in $G(\beta, \lambda)$ can be represented as $a\mu_+ + (1 - a)\mu_-$ with $0 < a < 1$, and a has the interpretation of fraction of mass of the liquid phase.

It is remarkable that it is possible to prove that the extremal states μ of $G(\beta, \lambda)$ enjoy the property of *ergodicity* in the sense that the above defined, see §5.4, corresponding dynamical systems (M, S, μ) are "ergodic" and they are the only points in (M, S, μ) with this property, see [Ru69].

The ergodicity property is the natural generalization of the notion introduced in the discrete evolution cases of the systems in Chap.I. We consider a family of commuting invertible transformations[7] $S = (S_1, \ldots, S_n)$ acting

---

[7] We shall only consider here and in the rest of the book *measurable* transformations, measurable functions, measurable sets. These are rather delicate notions, on the brink of the imponderable because to find nontrivial examples of *nonmeasurable* corresponding objects one needs the sinister axiom of choice. However if one wants to discuss notions like ergodicity in systems that are not regarded as discretized, abandoning Boltzmann's wise discrete conception of the world, one must say a few words on measurability. The spaces $M, M' \ldots$ that we consider here and later will all have a natural notion of "closeness" bewteen points, a *topology* in Mathematics: typically a metric can be defined on them (this metric can be defined but often it is not really useful so that it is not always explicitly defined as there is little doubt about what it could be). Therefore it makes sense to define *open sets*. One *declares all of them measurable*: more generally the smallest family of sets that contains all the open sets and that is closed under the operations of countable union, complementation and intersection is *by definition* the

on a space M so that setting $S^{\underline{k}} \equiv S_1^{k_1} \cdot S_2^{k_2} \cdot \ldots \cdot S_n^{k_n}$, with $\underline{k} = (k_1, \ldots, k_n)$ an n-ple of integers, we can define $S^{\underline{k}}x$ for $x \in M$. We say that a probability distribution μ on M is S-invariant if for every measurable set $E \subset M$ it is $\mu(S^{\underline{k}}E) = \mu(E)$. The triple (M, S, μ) is called a *discrete dynamical system* and $S^{\underline{k}}$ is called a "translation of $\underline{k}$ by S" (here S_j^0 is the identity map), [AA68]:

Definition (ergodicity): Let (M, S, μ) be a discrete dynamical system; it is "ergodic" if there are no nontrivial constants of motion, i.e. no measurable functions $x \to F(x)$ on a phase space M which are invariant under "translation by S" and which are not constant as x varies excluding, possibly, a set of zero μ-probability.

The definition can be extended in the obvious way to the case in which S is a *continuous flow, i.e.* $\underline{k} \in R^n$ and S_j^k are commuting transformations which satisfy the group property $S_j^k S_j^h = S_j^{h+k}$ for all $h, k \in R$ and S_j^0 is the identity map.

The above dynamical system with M being the phase space points of an infinite system, with *the "evolution" S being the spatial translations* and with μ being a Gibbs state, is ergodic if one cannot find observables that are *translation invariant* and at the same time not constant (outside a possible set of zero μ-probability).

If we sample the system configurations from an ergodic distribution μ we must find that the translation-invariant observables always have the same value. Thus for instance the global density (or specific volume) will always have the same value on all configurations that are sampled with distribution μ.

We see that because of the above-quoted theorem of unique decomposability of the elements of $G(\beta, \lambda)$ into extremal distributions the extremal points of $G(\beta, \lambda)$ deserve the name of *pure phases* as there is no way to see that they consist of different configurations by measuring global, translation-invariant, properties that they enjoy.

If, instead, a probability distribution in $G(\beta, \lambda)$ is not pure but it is a mixture of, say, two pure states with different densities and with coefficients

---

family of measurable sets, or *Borel sets*. A mesurable transformation S is a map of M into M' such that $S^{-1}E$ is measurable for any measurable E. Therefore it makes sense to say that a function (*i.e.* a map of M to R) is measurable. A probability distribution, or a "normalized measure", μ on M is a function defined on the measurable sets with values $\mu(E) \ge 0$ and which is additive (*i.e.* if $E = \cup_0^\infty E_n$ and the E_n's are pairwise disjoint, then $\mu(E) = \sum \mu(E_n)$), and such that $\mu(M) = 1$. Given a distribution μ on M one calls *μ-measurable* any set in the smallest collection of sets, closed under countable union, complementation and intersection, that contains the measurable sets *as well as any* other set that can be enclosed into a measurable set with 0 μ-measure: the latter are called "0 μ-measure sets". Likewise we can define μ-measurable functions and μ-measurable maps. The above notion of μ-measurability should not be confused with the previous notion of measurability. Why the name "measurable"? because a measurable function of one veriable is the most general function for which it is possible to set up, *in principle*, a table of values, *i.e.* a function that can be approximated by piecewise constant functions.

a and $1-a$ then by sampling the system configurations we may get either a configuration of the dense phase or one of the rarefied phase; so that the global density, which is a translation-invariant observable, can have two distinct values (each with a probability of occurrence in samples given by a and $1-a$ respectively), so that it is not constant.

An important general theorem for dynamical systems is *Birkhoff's ergodic theorem*, [AA68]:

Theorem (Birkhoff): Let (M,S,μ) *be a discrete dynamical system and let* Λ *be a cube of side* L*. Then for all* μ*-measurable functions* f *on* M *the limit* $\lim_{L\to\infty}\frac{1}{|\Lambda|}\sum_{\underline{k}\in\Lambda} f(S^{\underline{k}}x) = \overline{f}(x)$ *exists apart, possibly, from a set of* x*'s of zero* μ*-probability. Hence if* (M,S,μ) *is ergodic* $\overline{f}(x)$ *is a constant for* μ*-almost all* x *and, therefore, it is equal to* $\int_M \mu(dy)f(y)$*.*

The above statement can be also formulated for the case in which S is a flow (*i.e.* $\underline{k}\in R^n$ is a continuous vector) and it is also a valid statement.

Another consequence of the ergodicity is that particles located in two far apart cubes are observed as if they were *independently distributed*, at least on the average over the boxes locations. This is also a property that intuitively should characterize the physically pure homogeneous phases.[8]

§5.6. Other Characterizations of Gibbs Distributions. The DLR Equations

Via the variational principle (5.5.13) one finds all the translation invariant Gibbs distributions, but on physical grounds, as remarked, we expect that there may also exist, under suitable circumstances, nontranslation-invariant Gibbs distributions; *i.e.*, with the notation of §5.4, in general we shall have that $G^0(\beta,\lambda)$ contains $G(\beta,\lambda)$, *but it does not coincide with* $G(\beta,\lambda)$.

Therefore it is useful to look also for other characterizations of Gibbs states which do not "discriminate" the nontranslation-invariant states. Such a

---

[8] A simple abstract argument proves the statement. Let $\mu\in G(\beta,\lambda)$ be ergodic and denote by $\rho(\Delta)$ the average over the configurations x of the number $N(x)$ of particles in the unit cube Δ with respect to the distribution μ; let $\rho(\Delta,\Delta')$ be the average of the product of the number of particles in the unit cube Δ times that in the unit cube Δ'. The translation invariance of μ implies that $\rho(\Delta)$ is independent of the location of the unit cube Δ and that $\rho(\Delta,\Delta')$ depends only on the relative position of the unit cubes Δ and Δ'. If Λ is a large volume paved by unit cubes Δ the average number of particles in Λ will be $\langle\frac{N}{|\Lambda|}\rangle = \frac{1}{|\Lambda|}\sum_{\Delta\subset\Lambda}\rho(\Delta)$ and the average $\langle\frac{N(N-1)}{|\Lambda|^2}\rangle$ will be $\frac{1}{|\Lambda|^2}\sum_{\Delta,\Delta'\subset\Lambda^2}\rho(\Delta,\Delta')$. Given a configuration x, the limit ρ as $\Lambda\to\infty$ of $\frac{N(x)}{|\Lambda|}$ will exist, possibly outside of a set of configurations with μ-probability 0: since this limit (when it exists) is obviously translation invariant as a function of x, it must be a constant (possibly outside a set of 0 probability, by Birkhoff's theorem above; nonconstancy would be against ergodicity). For the same reason also $\frac{N(x)(N(x)-1)}{|\Lambda|^2}$ will have a limit equal to that of $\frac{N(x)^2}{|\Lambda|^2}$ which has to be constant and therefore equal to ρ^2. Hence $\frac{1}{|\Lambda|^2}\sum_{\Delta,\Delta'}\rho(\Delta,\Delta') - (\frac{1}{|\Lambda|}\sum_{\Delta}\rho(\Delta))^2 \xrightarrow[\Lambda\to\infty]{} 0$ which means that $\rho(\Delta,\Delta')\simeq\rho(\Delta)\rho(\Delta')$ "on the average over Δ,Δ'".

characterization is possible and is suggested by a heuristic argument based on the finite volume grand canonical distribution $\mu^{(\beta,\lambda,V)}$ without external fixed particles, (5.3.2).

We ask: given $V^0 \subset V$ what is the probability of finding inside V^0 exactly g particles in the positions $\underline{q}_1, \ldots, \underline{q}_g$ knowing that out of V^0 the particles are located in the positions $\underline{q}'_1, \underline{q}'_2, \ldots$?

Denoting by $\overline{f}_{V^0}(\underline{q}_1, \ldots, \underline{q}_g \,|\, \underline{q}'_1, \underline{q}'_2, \ldots) = \overline{f}_{V^0_,g}(\underline{q}|\underline{q}')$ the density of this *conditional probability*, where $\underline{q}$ abbreviates $(\underline{q}_1, \ldots, \underline{q}_g)$ and $\underline{q}'$ abbreviates $(\underline{q}'_1, \underline{q}'_2, \ldots)$, then it is immediate to deduce from (5.3.2) that:

5.6.1
$$\overline{f}_{V^0,g}(\underline{q}|\underline{q}') = \frac{\gamma(\beta)^{3g} e^{\left[-\beta\lambda g - \beta\Phi(\underline{q}_1,\ldots\underline{q}_g) - \beta\sum_{i=1}^{g}\sum_{j\geq 1}\varphi(\underline{q}_i - \underline{q}'_j)\right]}}{\text{normalization}} \tag{5.6.1}$$

where $\gamma(\beta) = (2\pi m\beta^{-1})^{\frac{3}{2}g}$ (see (5.5.2)) and the normalization is determined by imposing the condition that $\overline{f}_{V_0}$ defines a probability distribution, *i.e.* that:

5.6.2
$$\sum_{g=0}^{\infty} \int \overline{f}_{V^0_,g}(\underline{q}|\underline{q}') \frac{d\underline{q}}{g!\gamma(\beta)^{3g}} = 1\,. \tag{5.6.2}$$

This relation depends on the total volume V *only* because $\underline{q}'_1, \underline{q}'_2 \ldots$, *i.e.* the particles of the configuration external to V^0, are constrained to be in V, *i.e.* in the global container of the system (and outside V^0).

It is therefore natural to define, as an alternative to §5.3, §5.4, a (infinite volume) Gibbs distribution on M with parameters (β, λ) as a distribution μ on M Maxwellian in the momenta and for which the probability for the event in which the particles inside a fixed finite volume V^0 are in $\underline{q}_1, \ldots, \underline{q}_g$, conditional to knowing that the particles outside the box V^0 are in $\underline{q}'_1, \underline{q}'_2 \ldots$ (with any momenta) is given by (5.6.1) without any restriction that the particles at $\underline{q}'_1, \underline{q}'_2, \ldots$ be inside a larger container V (because the latter has, now, to be thought as infinite).

This reading of (5.6.1) is known as the *DLR equation* and it was proposed as a very general definition of Gibbs state (in the thermodynamic limit) by Dobrushin, Lanford, Ruelle, [Do68], [LR69].

This is important because one can establish, quite generally, the theorem that Gibbs distributions, defined as the probability distributions on M which are Maxwellian in the velocities (with the same inverse temperature parameter β) and which satisfy (5.6.1), *coincide* with the distributions in $G^0(\beta, \lambda)$ defined via the thermodynamic limit in the previous sections, whether or not they are translation invariant.

This is a theorem that holds as stated in the case of hard core systems; its validity in more general situations still presents a few technical problems to be understood although various weak versions of it exist in most cases of interest (*e.g.* in the case of superstable potentials), [LP76].

§5.7. Gibbs Distributions and Stochastic Processes

By integration of the momentum coordinates $\underline{p}$, the probability distributions μ on the "infinite volume" phase space M define corresponding probability distributions on the space M_1 of the position configurations of infinitely many particles $\underline{q}$. Thus Gibbs distributions integrated over the momentum variables (which can be "disregarded" as playing a trivial mathematical role from the point of view of the description of the states, although they are physically very important), provide us with an interesting class of distributions on M_1 which we shall still simply call Gibbs distributions (rather than using a pedantic distinction between Gibbs distributions and configurational Gibbs distributions).

In general the probability distributions μ on M_1 are known in probability theory as stochastic *point processes* because a point $\underline{q} \in M_1$ in fact describes a family of points, *i.e.* particles located in $\underline{q}_1, \underline{q}_2, \ldots$ in R^3, if $\underline{q} = (\underline{q}_1, \underline{q}_2, \ldots)$.

The remark permits us to give a new physical interpretation to several results of the general theory of point stochastic processes, and mainly it induces a translation of problems relevant for physics into interesting mathematical problems in the theory of point stochastic processes.

The issue that is, perhaps, central is to show that there exist simple choices of the interparticle potentials φ, assumed with hard core for simplicity, and of the parameters β, λ for which the variational principle or the DLR equations admit *more than one* solution.

This is the same as the problem of the existence of *phase transitions* in a homogeneous system of identical particles: in fact we have argued that the physical pure phases that can coexist can be identified with the solutions of the variational principle or of the DLR equations.

We have seen above that the van der Waals theory provides us with an affirmative answer to this issue; however it is rather unsatisfactory and, to date, there is still no example that can be treated without uncontrolled approximations (*i.e.* without introducing *ad hoc* hypotheses at the "right moment"). The above nice thermodynamic formalism might be empty, after all: but this possibility is really remote, and it is certainly not realized in models that are somewhat simpler than the ones so far used for continuous gases: see Chap.VI and the recent breakthrough in [LMP98].

Other remarkable problems that arise in the theory of stochastic processes and, independently, in the theory of phase transitions are related to questions of scale invariance.

From experience and from the phenomenological theories of phase transitions not only does the hypothesis emerge that the liquid-gas transition really takes place whenever the interaction potential φ has, besides a repulsive core, an attractive tail, but also the hypothesis that such a transition has a *critical point* (λ_c, β_c) where the Gibbs distribution (and the corresponding stochastic process) μ has special scaling properties, [Fi98], [BG95].

More precisely imagine that we pave the ambient space R^3 with a lattice of cubes $Q^L_{\underline{n}}$, with side L and parameterized by three integers $\underline{n} = (n_1, n_2, n_3)$,

so that the cube $Q^L_{\underline{n}}$ consists of the points with coordinates $n_h L \le x_h < (n_h+1)L$, $h = 1,2,3$.

Define the family of variables (*i.e.* "functions" on phase space, in probability theory language, or "observables" in physics language) $\sigma_{\underline{n}}$ on M:

5.7.1 $$\sigma_{\underline{n}} = [(\text{particles number in } Q^L_{\underline{n}}) - v(\mu)^{-1}L^3]/L^{3\delta/2} \tag{5.7.1}$$

where δ is a parameter to be chosen.

One gets a stochastic process, *i.e.* a probability distribution on a *space of states* consisting of the sequences $\{\sigma_{\underline{n}}\}$ indexed by $\underline{n} \in Z^3$, in which the "states at the site $\underline{n}$" are real numbers labeled by $\underline{n}$ and defined by (5.7.1).[9]

By these phenomenological theories of the critical point, we may expect that, in the limit as $L \to \infty$ and if δ is suitably chosen, the stochastic process describing the distribution of the variables $\sigma_{\underline{n}}$ tends to a limiting process such that, in the limit, the $\sigma_{\underline{n}}$ can be represented as:

5.7.2 $$\sigma_{\underline{n}} = \int_{Q^L_{\underline{n}}} \psi(x)\,\mathrm{d}x \tag{5.7.2}$$

where $\psi(x)$ are random variables: this is a "stochastic process on R^3" with *homogeneous correlation functions*; *i.e.* for every k and $x_1, x_2, \ldots, x_k$:

5.7.3 $$\langle\psi(x_1)\ldots\psi(x_k))\rangle = \text{ homogeneous function of } (x_1,\ldots,x_k) \tag{5.7.3}$$

if $\langle\cdot\rangle$ denotes the operation of evaluation of the average value (also called "expectation" in Probability Theory) with respect to the distribution of the random variables ψ.

Since no nontrivial examples of point stochastic processes with the above properties are known (or, better, were known until recently) one understands the interest, even from a purely mathematical viewpoint, of the theory of phase transitions which in its heuristic aspects provides a solution to various problems related to the existence and structure of stochastic processes. The heuristic results suggest in fact very challenging mathematical conjectures, and some ideas for their understanding (often only partial), so that the subject continues to attract the attention of many, [Wi83], [WF72], [Ga76], [Fi98].

It seems fair to say that the tumultuous development of statistical mechanics and of the theory of phase transitions has literally revolutionized the theory of probability as well.

We conclude by mentioning (we come back on this point later) that so far we have only discussed the properties of the Gibbs states as equilibrium states, but *without ever introducing the dynamics.* We have regarded them

---

[9] More generally one calls *stochastic process* a probability distribution on a space of states consisting of families of variables, called *random variables*, indexed by an arbitrary label; and both the label and the labeled variables can be in any space, a *ghastly generality.*

as dynamical systems with respect to *space* translations and we have seen that this leads to the mathematical definition of pure phase. One may wonder whether one could obtain similarly interesting notions by regarding the Gibbs states as stationary states for time evolution as well (*i.e.* for translations in time). This is a much harder question and we defer discussing it to Chap.IX.

§5.8. Absence of Phase Transitions: $d = 1$. Symmetries: $d = 2$

(A) One dimension. After the above general analysis and after setting up a formalism well suited for our programs we return to more concrete questions. We begin by showing that, as already stated several times, one dimensional systems with finite range interactions cannot have phase transitions of any sort, unless it is considered in the somewhat unphysical situation of having zero absolute temperature. Again we limit ourselves to the simple case of hard core interactions and call r_0 the hard core size (so that $\varphi(r) = +\infty$ if $r \le r_0$).

We shall use here as a definition of phase transition the presence of a horizontal segment in the graph of the pressure as a function of the specific volume at constant temperature. But other definitions could be used, *e.g.* the inequivalence of some of the ensembles and the dependence of the thermodynamic limit on the boundary conditions, discussed in §2.5 and above, in the present chapter.

Consider first the case in which the potential vanishes beyond $r = 2r_0$ where r_0 is the hard core radius: this case is particularly easy and is called the "*nearest neighbor*" interaction case.

It is best to use the pressure ensemble, see §2.5, (2.5.6), (2.5.16), with the volume V taking the continuum of values between 0 and ∞.

Then the partition function in the pressure ensemble, see (2.5.17), (2.2.1), is

$$J_N(\beta, p) = \ell(\beta)^{-N} \int_0^\infty \frac{dL}{L_0} \int_{[0,L]} \frac{dq_1 \dots dq_N}{N!} \cdot e^{-\beta p L} \cdot \prod_{j=1}^{N-1} e^{-\beta\varphi(q_{j+1}-q_j)} \tag{5.8.1}$$

5.8.1

where $\ell(\beta)^{-1} \overset{def}{=} \sqrt{\frac{2\pi m}{\beta h^2}}$.

We shall use the fact that the interaction cannot extend beyond the nearest neighbor and we label the particles $1, \dots, N$ so that $q_1 < q_2 < \dots < q_N$. In this way we restrict the integration domain by a factor $N!$. Thus, extending the integral to the region $0 < q_1 < q_2 < \dots < q_N < L$ we get rid of the $N!^{-1}$ present in the definition of the partition function.

The momentum integration yields the square root in front of the integral (it is raised to the power 1 rather than the usual 3 because the space dimension is now 1). The length L_0 is an arbitrary dimensional factor (see (2.5.17)).

Then we note that $L = q_1 + \big(\sum_{j=1}^{N-1}(q_{j+1} - q_j)\big) + L - q_N$ and introducing

the variables $q_{j+1} - q_j$ as independent variables, we see that:

5.8.2
$$J_N(\beta,p) = \ell(\beta)^{-N} \frac{1}{(\beta p)^2 \, L_0} \Big(\int_0^\infty e^{-\beta p q} e^{-\beta\varphi(q)} dq \Big)^{N-1} \tag{5.8.2}$$

so that the thermodynamic limit $\lim_{N\to\infty} \frac{1}{N} \log J_N(\beta,p)$ is

5.8.3
$$-\beta\lambda(\beta,p) = \log \Big(\sqrt{\frac{2\pi m}{\beta h^2}} \cdot \int_0^\infty e^{-\beta p q} e^{-\beta\varphi(q)} \, dq \Big) . \tag{5.8.3}$$

Equivalence between pressure ensemble and canonical ensemble is worked out along the same lines in which in §2.5 equivalence between canonical and grand canonical ensembles (hence orthodicity) were derived.

One finds, as mentioned in §2.5, that the quantity p can be identified with pressure and $\lambda(\beta,p)$ can be identified with the Gibbs potential $u - Ts + pv$ (see (2.5.11)), and $\beta = 1/k_B T$. Moreover the equation of state is derived (by using the orthodicity) from the thermodynamic relation $(\frac{\partial\lambda}{\partial p})_\beta = v$.

Relation (5.8.3) implies that the Gibbs potential $\lambda(\beta,p)$ is analytic in β, p for $\beta, p > 0$; and it is strictly monotonic in p so that the relation $(\frac{\partial\lambda}{\partial p})_\beta = v$ implies that pressure is analytic and strictly monotonic (decreasing) in v: hence the equation of state cannot have any phase transition plateau.

The above analysis is a special case of *Van Hove's theorem*, which holds for interactions extending beyond the nearest neighbor, see Appendix 5.A1 below, and it played an important role in making clear that short-range one-dimensional systems cannot undergo phase transitions, [VH50]. Further extensions can be found in [GMR69].

If one adopts the definition of phase transition based on sensitivity of the thermodynamic limit to variations of boundary conditions one can give a more general, conceptually simpler, argument to show that in one-dimensional systems there cannot be any phase transition if the potential energy of mutual interaction between a configuration $\underline{q}$ of particles to the left of a reference particle (located at the origin O, say) and one configuration $\underline{q}'$ to the right of the particle (with $\underline{q} \cup O \cup \underline{q}'$ compatible with the hard cores) is uniformly bounded.

The argument, due to Landau, is simply that, in this case, the distribution of the configurations to the right of a point and to the left of it are essentially independent: hence by changing the configuration of fixed particles outside a box one does not alter appreciably the probability distribution inside it.

This is so because the weight of a configuration $\underline{q}$, consisting of a part $\underline{q}_1$ to the left of the origin and of a part $\underline{q}_2$ to the right of it, is the exponential of $-\beta H$, if H is the energy of the configuration. But the energy of such a configuration is a sum of two quantities (large, of the order of the volume occupied by the configurations) which are the sum of the energies that each of the two parts $\underline{q}_1$ and $\underline{q}_2$ would have, in the absence of the other part, plus the mutual energy. The latter is, however, *bounded independently* of the choice of $\underline{q}_1$ and $\underline{q}_2$.

In itself this does not immediately imply that there can be no dependence on boundary conditions because a finite ratio between two probabilities is not the same thing as a ratio close to 1: hence the argument has to be refined by considerations that show that it also implies actual boundary condition independence: I find it easier to just give analytic details on the above argument (see Appendix 5.A1 below) rather than indulging on heuristic discussions. The argument clearly shows a mechanism responsible for the "loss" of memory of the boundary conditions as one proceeds in from the boundary down to the center of a finite interval $[0, L]$.

Hence the larger the box the smaller is the influence of the external particles on the bulk of the particles in the box: hence no inequivalence between the ensembles can arise, *i.e.* no phase transitions in that sense. One also says that *no long-range order can be established in such systems*, in the sense that one loses memory of the boundary conditions as the boundaries recede to infinity in the process of taking the thermodynamic limit.

Note that the argument above fails if the space dimension is ≥ 2: in this case *even if the interaction is short ranged* the energy of interaction between two regions of space separated by a boundary is *of the order of the boundary area*. Hence one cannot bound above and below the probability of any two configurations in two half-spaces by the product of the probabilities of the two configurations, each computed as if the other was not there (because the bound would be proportional to the exponential of the surface of separation, which tends to ∞ when the surface grows large). This means that we cannot consider, at least not in general, the configurations in the two half spaces as independently distributed.

Analytically a condition sufficient to imply that the energy between a configuration to the left and one to the right of the origin is bounded above, if the dimesion d is $d = 1$, is simply expressed (as it is easy to check) by:

5.8.4
$$\int_{2r_0}^{\infty} r \mid \varphi(r) \mid dr < +\infty \tag{5.8.4}$$

One usually says, therefore, that in order to have phase transitions in $d = 1$ systems one needs a potential that is "so long range" that it has *divergent first moment*. It can be shown by counterexamples that if the condition (5.8.4) fails there can be phase transitions even in one-dimensional systems, at least in further simplified models, [Dy69]. In fact very recently the first phase transition in a continuous system and *in the absence of symmetry breaking* has been proved to occur precisely in a system violating (5.8.4), see [Jo95].

The arguments of this section apply also to discrete models like *lattice gases* or *lattice spin* models, see §6.2 below and Chap.VI, Chap.VII, and §9.7.

(B) Symmetries. By *symmetry* one means a group of transformations acting on the configurations of a system subject to some boundary condition (*e.g.* periodic or open) and transforming each of them into configurations with the same energy and with the same boundary condition.

Systems with "too much symmetry" sometimes cannot show phase transitions. This is best discussed if one uses as definition of phase transition the existence of *long-range order*.

The latter is defined by considering a localized observable F (see (5.4.2)) which has zero average $\langle F \rangle$ in all Gibbs states obtained as thermodynamic limits, with suitable boundary conditions (*e.g.* periodic or open), because of their symmetry properties.

Suppose that the average value $\langle F \cdot \tau_\xi F \rangle$ of the product of the observable times a translate[10] by ξ of itself, a quantity called the spatial *autocorrelation* of the observable, *does not approach* 0 as $\xi \to \infty$. Then one says that the system shows *long-range order* for the *order parameter* F.

The symmetry is *continuous* if the group of transformations is a continuous group. For instance continuous systems have translational symmetry if considered with periodic boundary conditions, so that the number n_Δ of particles in a small box Δ is a local observable and such is also $\nu_\Delta \stackrel{def}{=} (n_\Delta - \overline{n}_\Delta)$ where $\overline{n}_\Delta$ is the average over translations. For symmetry reasons this quantity has zero average in the Gibbs states associated with the Hamiltonian describing the system with the symmetric boundary condition (periodic in this case). Denote by $\Delta + \xi$ the box Δ' obtained by translating Δ by a vector ξ and let $\langle \cdot \rangle$ denote the average in one element of an orthodic ensemble (*i.e.* an average with respect to a *Gibbs state*). The system is said to *show long-range order* if the autocorrelation function at distance ξ is a function $\langle \nu_\Delta \nu_{\Delta+\xi} \rangle$ which *does not tend to zero as* $\xi \to \infty$.

Note that failure of convergence to zero of $\langle \nu_\Delta \nu_{\Delta+\xi} \rangle$, see also §5.5, footnote 8, is precisely what we expect should happen if the system had a *crystalline phase* (in which case the $\langle \nu_\Delta \nu_{\Delta+\xi} \rangle$ should show an oscillatory behavior, in ξ, of the correlation function). One can also prove that long range order of some observable implies that the derivative of the pressure (or of other thermodynamic functions) with respect to suitable perturbations of the energy function has a discontinuity, so there is an intimate connection between phase transitions defined in terms of long-range order and in terms of singularities of thermodynamic functions.

As an example of an application of a general theorem, the *Mermin-Wagner theorem*, [MW66], [Me67],[Ru69], one can state that *if the dimension of the ambient space is* $d = 2$ then a system which in periodic boundary conditions shows a *continuous symmetry* cannot have any local observable whose average vanishes and whose autocorrelations at distance ξ *do not tend to zero* as the distance $\xi \to \infty$. This theorem is the first of a series of similar theorems based on an important kind of inequality called the *infrared inequality* and it has led to developments that solved several long-standing problems, see for instance [Fr81], [DLS78]. Here we choose not to enter into more details in spite of the great importance of the technique and the reader is referred

---

[10] A translate by the vector ξ of an observable F is defined as the observable $\tau_\xi F$ such that $\tau_\xi F(x) = F(x + \xi)$ where $x + \xi$ is the configuration obtained from x by translating by ξ all particles positions, leaving the velocities unchanged.

to the literature.

The limitation to dimension $d = 2$ is, however, a strong limitation to the generality of the theorem and very seldom does it apply to higher-dimensional systems. More precisely systems can be divided into classes each of which has a "critical dimension" below which too much symmetry implies the absence of phase transitions (or of certain kinds of phase transitions), see [WF72], [Fr81], [Fr86], [Fi98].

§5.9. Absence of Phase Transitions: High Temperature and the KS Equations

There is *another* class of systems in which no phase transitions take place. These are the systems so far considered (with stable and tempered interactions, see §2.2) in states with high temperature and low density.

We use here as definition of phase transition that of a singularity in the equation of state, although in the cases below one could show that phase transitions do not occur even in other senses (like persisting sensitivity to boundary conditions as the boundaries recede to ∞).

One can easily show the absence of phase transitions for β^{-1} and v large by showing that the equation of state is analytic. In fact in such regions the virial series, (5.1.16), is convergent and we have analyticity in v^{-1} and β of the equation of state.

There are two ways of attacking the problem: one is rather direct and looks for an algorithm that constructs the coefficients of the virial series. The algorithm can be found quite easily: but the k-th order term results as a sum of very many terms (a number growing more than exponentially fast in the order k) and it is not so easy (although it can be done) to show by combinatorial arguments that their sum is bounded by $c(\beta)^k$ if β is small enough, [Gr62],[Pe63], see also equation (4.2), p. 176 in [GMM72], dealing with a case only apparently different and in fact more general.

The other approach is somewhat less natural but it leads quite easily to the desired solution. It attempts to solve a much more general question. Namely the problem of computing the functions $\overline{f}_{V^0}$ of §5.5 and (5.5.1), *i.e.* "*all the properties*" of the system.

We consider a gas in a cubic container V and with an interaction potential φ satisfying (4.1.1). The state of the system, in the grand canonical ensemble, can be defined in terms of the local distributions discussed in §5.4, (5.4.1), or in terms of the more convenient ("spatial or configurational") *correlation functions*

5.9.1
$$\rho_V(q_1,\ldots,q_n) = \\ = \frac{1}{\Xi_V(\beta,\lambda)}\sum_{m=0}^{\infty} z^{n+m}\int e^{-\beta\Phi(q_1,\ldots,q_n,y_1,\ldots,y_m)}\,\frac{dy_1\ldots dy_m}{m!} \tag{5.9.1}$$

where $z = e^{-\beta\lambda}(\sqrt{2\pi m\beta^{-1}h^{-2}})^3$ is called the *activity*: it has the dimension of a density *i.e.* of a length^{-3} as we included in it also the factor h^{-3m} which

in (5.5.1) was included in $\overline{f}_{V_0}$. The correlation functions are, therefore, the probability densities for finding n particles at the positions $q_1, \ldots, q_n$ *with any momenta and irrespective of where the other particles are*. The square root comes from the integration over the momenta variables (which drop out of the scene, with no regret as they play a trivial role in classical statistical mechanics). The integral over the y's is over the volume V.

The energy $\Phi(q_1, \ldots, q_n, y_1, \ldots, y_m)$ can be decomposed as:

5.9.2
$$\Phi_1(q_1; q_2, \ldots, q_n) + \sum_{j=1}^{m} \varphi(q_1 - y_j) + \Phi(q_2, \ldots, q_n, y_1, \ldots, y_m) \qquad (5.9.2)$$

where $\Phi_1(q_1; q_2, \ldots, q_n) = \sum_{i=2}^{n} \varphi(q_1 - q_i)$ is the energy of interaction of particle q_1 with the group of $(q_2, \ldots, q_n)$.

We can imagine that q_1 is the "most interacting particle" among the $(q_1, \ldots, q_n)$, *i.e.* it is one that maximizes the potential energy of interaction with the group of the other particles. Since $\Phi(q_1, \ldots, q_n) \geq -Bn$ by stability, this implies that

5.9.3
$$\Phi_1(q_1; q_2, \ldots, q_n) \geq -2B\,. \qquad (5.9.3)$$

In some special cases the selection of q_1 among the n particles $q_1, \ldots, q_n$ may be ambiguous; the choice then can be made arbitrarily (for the purposes of the following argument).

Then from the definition (5.9.1) and by using the decomposition (5.9.2) and if Ξ_V denotes the grand canonical partition function in the volume V, we see that we have the following simple algebraic identities

5.9.4
$$\rho_V(q_1, \ldots, q_n) = \frac{1}{\Xi_V}\, z e^{-\beta \Phi_1(q_1; q_2, \ldots, q_n)} \sum_{m=0}^{\infty} z^{n-1+m} \cdot \qquad (5.9.4)$$
$$\cdot \int e^{-\beta \sum_j \varphi(q_1 - y_j)} e^{-\beta \Phi(q_2, \ldots, q_n, y_1, \ldots, y_m)} \frac{dy_1 \ldots dy_m}{m!} =$$
$$= \frac{1}{\Xi_V} z e^{-\beta \Phi_1(q_1; q_2, \ldots, q_n)} \sum_{m=0}^{\infty} z^{n-1+m} \cdot$$
$$\cdot \int \Big[\prod_{j=1}^{m} \Big(1 + (e^{-\beta \varphi(q_1 - y_j)} - 1) \Big) \Big] e^{-\beta \Phi(q_2, \ldots, q_n, y_1, \ldots, y_m)} \frac{dy_1 \ldots dy_m}{m!} =$$
$$= \frac{1}{\Xi_V} z e^{-\beta \Phi_1(q_1; q_2, \ldots, q_n)} \sum_{m=0}^{\infty} \sum_{s=0}^{m} z^{n-1+m} \sum_{j_1, j_2, \ldots, j_s;\, 1 \leq j_k \leq m} \cdot$$
$$\cdot \int \frac{dy_1 \ldots dy_m}{m!} \prod_{k=1}^{s} \Big(e^{-\beta \varphi(q_1 - y_{j_k})} - 1 \Big) e^{-\beta \Phi(q_2, \ldots, q_n, y_1, \ldots, y_m)}$$

having developed the product in the second line. By using the symmetry in the y_j variables we can suppose that the $j_1, \ldots, j_s$ are in fact $1, 2, \ldots, s$

and rewrite (5.9.4) as

$$\rho_V(q_1,\ldots,q_n) = = \frac{1}{\Xi_V}\, z e^{-\beta\Phi_1(q_1;q_2,\ldots,q_n)} \sum_{s=0}^{\infty} z^s \int \frac{dy_1\ldots dy_s}{s!}\cdot \prod_{k=1}^{s}\Big(e^{-\beta\varphi(q_1-y_k)}-1\Big)\sum_{t=0}^{\infty} z^{n-1+t}\int\frac{dy'_1\ldots dy'_t}{t!}e^{-\beta\Phi(q_2,\ldots,y_1,\ldots,y_s,y'_1,\ldots y'_t)} \tag{5.9.5}$$

5.9.5 having in the last step called $m = s + t$ and replaced, with the appropriate combinatorial factors required by the change, $(y_1,\ldots,y_m)$ with $(y_1,\ldots,y_s,y'_1,\ldots,y'_t)$. Hence we see that the integrals reconstruct the correlation functions and (5.9.4) becomes

5.9.6
$$\rho_V(q_1,\ldots,q_n) = z e^{-\beta\Phi_1(q_1;q_2,\ldots,q_n)} \sum_{s=0}^{\infty}\int \frac{dy_1\ldots dy_s}{s!}\cdot \cdot\prod_{k=0}^{s}\Big(e^{-\beta\varphi(q_1-y_k)}-1\Big)\rho_V(q_2,\ldots,q_n,y_1,\ldots,y_s) \tag{5.9.6}$$

in which the term with $s = 0$ has to be interpreted as:

5.9.7
$$\begin{array}{ll} z\, e^{-\beta\Phi_1(q_1;q_2,\ldots,q_n)}\,\rho_V(q_2,\ldots,q_n), & \text{if } n>1\\ z & \text{if } n=1\end{array} \tag{5.9.7}$$

and all the variables q, y are considered to be in V.

Relations (5.9.6) are called the *Kirkwood-Salsburg equations*: they are important because we can use them to show that the virial series converges for β and v small. And in fact they allow us to obtain a *complete theory* of the gases in such regimes of β, v.

We can regard ρ_V as a *sequence* of functions "of one, two,... particle positions": $\rho_V = \{\rho_V(q_1,\ldots,q_n)\}_{n;\,q_1,\ldots}$ vanishing for $q_j \notin V$. If we define the sequence α_V of functions of one, two,... particle positions by setting $\alpha_V(q_1) \equiv 1$ if $q_1 \in V$ and $\alpha_V(q_1,\ldots,q_n) \equiv 0$ if $n > 1$ or $n = 1, q_j \notin V$, then we can write (5.9.7) as

5.9.8
$$\rho_V = z\alpha_V + zK\rho_V \tag{5.9.8}$$

where, if $\delta_{n>1} = 0$ for $n = 1$ and $\delta_{n>1} = 1$ for $n > 1$,

5.9.9
$$K\,\rho_V(q_1,\ldots,q_n) = e^{-\beta\Phi_1(q_1;q_2,\ldots,q_n)}\Big(\rho_V(q_2,\ldots,q_n)\,\delta_{n>1}+ +\sum_{s=1}^{\infty}\int\frac{dy_1\ldots dy_s}{s!}\prod_{k=1}^{s}\big(e^{-\beta\varphi(q_1-y_k)}-1\big)\rho_V(q_2,\ldots,q_n,y_1,\ldots,y_s)\Big) \tag{5.9.9}$$

which shows that the Kirkwood-Salsburg equations can be regarded as *linear inhomogeneous* "integral" equations for the family of correlation functions

that describe a given system in the box V. The kernel K of these equations is *independent of* V.

We should note that the quantities in ρ_V have different physical dimensions. In fact $\rho_V(q_1,\ldots,q_n)$ has the dimension of a length to the power $-3n$.

This sounds *bad enough* to wish to write (5.9.9) in *dimensionless* form. For this we need a length scale, and a natural choice is the "range" of the potential that could be defined as $\frac{\int|\varphi(x)||x|dx}{\int|\varphi(x)|dx}$. But a more convenient length that we can associate with our system is the quantity $r(\beta)$:

5.9.10
$$r(\beta)^3 \equiv \int_{R^3} |e^{-\beta\varphi(q)} - 1| d^3q \tag{5.9.10}$$

which can be called the *effective range* at inverse temperature β. Note that $r(\beta) \xrightarrow[\beta\to 0]{} 0$ if the potential has no hard core; if the potential has a hard core with radius a, and it is smooth and bounded otherwise, then $r(\beta)^3 \xrightarrow[\beta\to 0]{} \frac{4\pi}{3}a^3$.

The length $r(\beta)$ can be used to define the *dimensionless correlations* $\overline{\rho}_V(q_1,\ldots,q_n)$ as:

5.9.11
$$\overline{\rho}_V(q_1,\ldots,q_n) = r(\beta)^{3n}\rho_V(q_1,\ldots,q_n) \tag{5.9.11}$$

and setting $\zeta = zr(\beta)^3$, the above equations can be written in dimensionless form:

5.9.12
$$\overline{\rho}_V = \zeta\alpha_V + \zeta\,\overline{K}\,\overline{\rho}_V, \tag{5.9.12}$$

with

5.9.13
$$\begin{aligned}\overline{K}\,\overline{\rho}_V(q_1,\ldots,q_n) &= e^{-\beta\Phi_1(q_1;q_2,\ldots,q_n)}\big(\overline{\rho}_V(q_2,\ldots,q_n) + \sum_{s=1}^{\infty}\cdot\\ &\cdot\int\frac{dy_1\ldots dy_s}{s!\,r(\beta)^{3s}}\prod_{k=0}^{s}\Big(e^{-\beta\varphi(q_1-y_k)}-1\Big)\,\overline{\rho}_V(q_2,\ldots,q_n,y_1,\ldots,y_s)\big)\,.\end{aligned} \tag{5.9.13}$$

Then we can write the recursive formula:

5.9.14
$$\overline{\rho}_V = \zeta\alpha_V + \zeta^2\,\overline{K}\alpha_V + \zeta^3\,\overline{K}^2\alpha_V + \zeta^4\,\overline{K}^3\alpha_V + \ldots \tag{5.9.14}$$

which gives us an expression for the correlation functions, *provided the series converges*, of course.

The convergence of the series is easily discussed if one notes that

5.9.15
$$\begin{aligned}|\,\overline{K}^p\alpha_V(q_1,\ldots,q_n)| &\le e^{2\beta B}\sum_{s=0}^{\infty}\int_{R^{3s}}\frac{dy_1\ldots dy_s}{s!\,r(\beta)^{3s}}\,\cdot\\ &\cdot\prod_{k=1}^{s}|e^{-\beta\varphi(q_1-y_k)}-1|\,|\,\overline{K}^{p-1}\alpha_V(q_2,\ldots,q_n,y_1,\ldots,y_s)|\end{aligned} \tag{5.9.15}$$

so that if we call $M(p) = \max_{n;\,q_1,\ldots,q_n} |K^p\alpha_V(q_1,\ldots,q_n)|$ we see that

5.9.16
$$M(p) \le e^{2\beta B} M(p-1)\big(1+\sum_{s=1}^{\infty} s!^{-1}\big) = e^{2\beta B+1} M(p-1) \qquad (5.9.16)$$

and $M(0)$ has to be set equal to 1. This implies that $M(p) \le e^{(2\beta B+1)p}$ so that the series (5.9.14) converges if $|z| < e^{-(2\beta B+1)} r(\beta)^{-3}$.

The convergence is uniform (as $V\to\infty$) and $(\overline{K}^p)\alpha_V(q_1,\ldots,q_n)$ tends to a limit as $V\to\infty$ at fixed $q_1,\ldots,q_n$ and the limit is simply $(\overline{K}^p\alpha)(q_1,\ldots,q_n)$ if $\alpha(q_1,\ldots,q_n)=0$ unless $n=1$, and $\alpha(q_1)\equiv 1$. This is because the kernel $\overline{K}$ contains the factors $(e^{-\beta\varphi(y-q_1)}-1)$ which will tend to zero for $y\to\infty$ not slower than $|\varphi(y)|$, *i.e.* summably by the temperedness condition. It is also clear that $(\overline{K}^p\alpha)(q_1,\ldots,q_n)$ is translation invariant.

Hence the limits as $V\to\infty$ of the correlation functions do exist and they can be computed by a convergent power series in z, and the correlation functions will be translation invariant in the thermodynamic limit and the lack of translation symmetry, due to the confinement in the box V, disappears when the box recedes to ∞.

In particular the one-point correlation function $\rho=\rho(q)$ is simply $\rho = z\big(1+O(zr(\beta)^3)\big)$, which to lowest order in z just shows that the activity can be identified with the density. *Activity and density essentially coincide when they are small.*

Furthermore $\beta\, p_V \equiv \frac{1}{V}\log\Xi_V(\beta,\mu)$ has the property that $(z\partial_z\beta p_V)_\beta = \frac{1}{V}\int\rho_V(q)dq$, as is immediately checked (by using the definition of ρ_V in (5.9.1)). Therefore the above remarks imply:

5.9.17
$$\beta p_V(\beta,z) = \lim_{V\to\infty}\frac{1}{V}\log\Xi(\beta,\lambda) = \int_0^z \frac{dz'}{z'}\rho(\beta,z') \qquad (5.9.17)$$

hence the grand canonical pressure $p(\beta,z)$ is analytic in β,z. The density ρ is analytic in z as well and $\rho\simeq z$ for z small. It follows that the *pressure is analytic in the density* and $\beta\, p = \rho\,(1+O(\rho^2))$, at small density. In other words the equation of state is, to lowest order, essentially the equation of a perfect gas, and all the quantities that we may want to study are analytic functions of temperature and density.

The system is essentially a free gas and it has no phase transitions in the sense of a discontinuity or a singularity in the dependence of a thermodynamic function in terms of others.

However the system also cannot show phase transitions in the sense of sensitive dependence on the boundary conditions: this is essentially clear from the above analysis (*i.e.* from the remarked short range nature of the kernel K) which shows that the dependence on the boundary condition disappears as the boundary recedes to infinity while translation invariance is recovered.

One could, nevertheless, think that by taking other boundary conditions the argument may fail. It can however be shown that this is not the case,

simply by pushing the above analysis a little further. The key remark is in fact that *any* infinite volume state, obtained by *any* sequence of boundary conditions with fixed external particles, will obey the DLR equations, see §5.6, and the latter *can be shown* to imply the "infinite volume Kirkwood-Salsburg equations, [La70]. The latter are simply (5.9.13) with V replaced by R^3, and make sense also for infinite volume as soon as the correlation functions satisfy a bound like $\rho(q_1, \ldots, q_n) \le \xi^n$ for some ξ no matter how large.

The limits of finite volume states with fixed external particle boundary conditions do satisfy a bound of this type and also the DLR equations and, therefore, the Kirkwood-Salsburg equations in infinite volume. The uniqueness of the solutions of such equations proves, in the region $|z|\, e^{2\beta B+1} r(\beta)^3 < 1$, the boundary condition independence (hence the translation invariance) of the Gibbs states, [Do68c],[LR69].

Finally one can also see that the state of the system can be regarded as describing a distribution of particles in which particles occupying regions that are far apart are "*independently distributed*". There are several ways to express this property. The simplest is to say that the correlations have a *cluster property*, see footnote 8, §5.5. This means that

5.9.18 $$\lim_{a\to\infty} \rho(q_1, \ldots, q_n, q'_1 + a, \ldots, q'_{n'} + a) = \rho(q_1, \ldots, q_n)\rho(q'_1, \ldots, q'_{n'}) \quad (5.9.18)$$

and this property is an immediate consequence of the above analysis in the small β, small ρ regions.

In fact, restricting ourselves for simplicity to the case in which the potential has finite range r_0 we easily check that

5.9.19 $$\begin{aligned} &z(zK)^p \alpha(q_1, \ldots, q_n, q'_1 + a, \ldots, q'_{n'} + a) = \\ &= \sum_{p_1+p_2=p} z(zK)^{p_1}(q_1, \ldots, q_n) z(zK)^{p_2}(q_1, \ldots, q_n) \end{aligned} \quad (5.9.19)$$

for all p and *provided the distance* between the cluster $q_1, \ldots, q_n$ and the cluster $q'_1 + a, \ldots, q'_{n'} + a$ is greater than $p\, r_0$.

This is satisfied by induction and implies that the power series expansion for the difference between the expression under the limit sign in (5.9.18) and the right-hand side *starts* at $p = O(a/r_0)$ because all the coefficients with $p < \frac{a}{r_0}$ must vanish due to the fact that the kernel of the operator K vanishes when its arguments contain points that are too far away. In fact if the argument contains $s+1$ points $q_1, y_1, \ldots, y_s$ and the maximum distance between them is greater than $(s+1)r_0$ then at least one is further away than r_0 from q_1, see (5.9.9). Then the above proved convergence shows that the limit *is approached exponentially at a speed that is at least* $(zr(\beta)^3 e^{2\beta B+1})^{a/r_0}$ (this being the rate of approach to zero of the remainder of a geometric series with ratio $zr(\beta)^3 e^{2\beta B+1}$ and starting at order $|a|/r_0$).

Hence if one wants to look for phase transitions *one must forget* the regions of low density and high temperature. The Kirkwood-Salsburg equations

are only one example of equations leading to convergent expansions for the correlation functions: there are many recent developments based on similar equations that are derived for other models or even for the same ones considered above. The most interesting concern *lattice models*. See §5.10 below: see [Ca83], [KP86], [Br86] for some examples.

§5.10. Phase Transitions and Models

As already mentioned the problem of showing the existence of phase transitions in models of homogeneous gases, which we have been considering so far, is in fact still open.

Therefore it makes sense to study the phase transitions problem in simpler models, tractable to some extent but nontrivial. In fact such an investigation can give a very detailed and deep understanding of the phase transition phenomenon.

The simplest models are the so-called *lattice models*. They are models in which the particles are constrained to occupy points of a lattice in space. In such models particles cannot move in the ordinary sense of the word (because they are on a lattice and the motion would have to take place by jumps) and therefore their configurations do not contain momenta variables.

The energy of interaction is just a potential energy and the ensembles are defined as probability distributions on the position coordinates of the particle configurations. Usually the potential is a *pair potential* decaying fast at ∞ and, often, with a hard core forbidding double or higher occupancy of the same lattice site.

Often the models allow at most one particle to occupy each lattice site. For instance the *nearest neighbor lattice gas*, on a square lattice with mesh $a>0$, is defined by the potential energy that is attributed to the configuration X of occupied sites:

5.10.1
$$H(X)=\sum_{x,y\in X}\varphi(x-y),\qquad \varphi(x)=\begin{cases} J & \text{if } |x|=a\\ 0 & \text{otherwise}.\end{cases}\tag{5.10.1}$$

One can define the *canonical ensemble*, with parameters β, N, in a box Λ simply as the probability distribution of the subsets of Λ with N points:

5.10.2
$$p(X)=\frac{e^{-\beta H(X)}}{\sum_{\substack{|X'|=N\\ X'\subset\Lambda}} e^{-\beta H(X')}}\qquad |X|=N\tag{5.10.2}$$

where $|X|$ is the number of points in the set X; and likewise the grand canonical ensemble with parameters β,λ in the box Λ by

5.10.3
$$p(X)=\frac{e^{-\beta\lambda|X|}e^{-\beta H(X)}}{\sum_{X'\subset\Lambda} e^{-\beta\lambda|X'|}e^{-\beta H(X')}}\,.\tag{5.10.3}$$

Finally we can remark that a lattice gas in which in each site there can be at most one particle can be regarded as a model for the distribution of a family of *spins* on a lattice.

Such models are quite common and useful: for instance they arise in studying systems with magnetic properties. One simply identifies as "occupied" a site with a "spin up" or + and as an "empty" site a site with a "spin down" or − (of course one could make the opposite choice). If $\underline{\sigma} = \{\underline{\sigma}_x\}_\Lambda$ is a spin configuration, the energy of the configuration will usually take the form

5.10.4
$$H(\underline{\sigma}) = \sum_{x,y\in\Lambda} \varphi(x-y)\sigma_x\sigma_y + h\sum_x \sigma_x \tag{5.10.4}$$

and one calls canonical and grand canonical ensembles in the box Λ with respective parameters β, M or β, h the probability distributions on the spin configurations $\underline{\sigma} = \{\sigma_x\}_\Lambda$ with $\sum_{x\in\Lambda}\sigma_x = M$ or without constraint on M, respectively, defined by

5.10.5
$$\begin{aligned} p_{\beta,M}(\underline{\sigma}) =& \frac{e^{-\beta\sum_{x,y}\varphi(x-y)\sigma_x\sigma_y}}{\sum e^{-\beta\sum_{x,y}\varphi(x-y)\sigma'_x\sigma'_y}} \\ p_{\beta,h}(\underline{\sigma}) =& \frac{e^{-\beta h\sum\sigma_x - \beta\sum_{x,y}\varphi(x-y)\sigma_x\sigma_y}}{\sum e^{-\beta h\sum\sigma'_x - \beta\sum_{x,y}\varphi(x-y)\sigma'_x\sigma'_y}} \end{aligned} \tag{5.10.5}$$

where the sums in the denominators run over the $\underline{\sigma}'$ with $\sum_x \sigma'_x = M$ in the first case and over all $\underline{\sigma}'$'s in the second case.

As in the study of the previous continuous systems one can define the canonical and grand canonical ensembles with "external fixed particle configurations" and the corresponding ensembles with "external fixed spin configurations".

For each configuration $X \subset \Lambda$ of a lattice gas we define $\{n_x\}$ to be $n_x = 1$ if $x \in X$ and $n_x = 0$ if $x \notin X$. Then the transformation:

5.10.6
$$\sigma_x = 2n_x - 1 \tag{5.10.6}$$

establishes a correspondence between lattice gas and spin distributions. In this correspondence lattice gases with canonical (or grand canonical) distributions and given boundary conditions with external fixed particles are mapped into canonical (or grand canonical) spin distributions with suitably correspondent boundary conditions of external fixed spins.

In the correspondence the potential $\varphi(x-y)$ of the lattice gas generates a potential $\frac{1}{4}\varphi(x-y)$ for the corresponding spin system. The chemical potential λ for the lattice gas becomes the magnetic field h for the spin system with $h = \frac{1}{2}(\lambda + \sum_{x\neq 0}\varphi(x))$:

5.10.7
$$\varphi(x) \to \varphi' = \frac{1}{4}\varphi(x); \qquad \lambda \to h = \frac{1}{2}(\lambda + \sum_{x\neq 0}\varphi(x))\,. \tag{5.10.7}$$

The correspondence between boundary conditions is also easy: for instance a boundary condition for the lattice gas in which all external sites are occupied

becomes a boundary condition in which all the external sites contain a spin $+$. The correspondence between lattice gas and spin systems is so complete that one often switches from one to the other with little discussion.

The thermodynamic limits for the partition functions of lattice gas models, defined by

5.10.8
$$\beta f(\beta,v) = -\lim_{\frac{|\Lambda|}{N}=v,\Lambda\to\infty} \frac{1}{N}\log\sum_{|X|=N,\,X\subset\Lambda} e^{-\beta H(X)} \tag{5.10.8}$$

and

5.10.9
$$\beta p(\beta,\lambda) = \lim_{\Lambda\to\infty}\frac{1}{\Lambda}\log\sum_{X\subset\Lambda} e^{-\beta H(X)-\beta\lambda|X|} \tag{5.10.9}$$

can be shown to exist, by an argument similar to that discussed in Chap.IV. (and by far easier). They have the same convexity and continuity properties of the corresponding quantities in the case of the continuous models and they will be given the same names (free energy and pressure). They are boundary condition independent, as was the case in the continuum models with hard core interactions.

Likewise the thermodynamic limits exist also for the spin models partition functions and they are denoted by f, p:

5.10.10
$$\beta f(\beta,m) = -\lim_{\Lambda\to\infty,\,\frac{M}{|\Lambda|}\to m}\frac{1}{\Lambda}\log\sum_{\sum_x\sigma_x=M} e^{-\beta H(\underline{\sigma})} \tag{5.10.10}$$

and

5.10.11
$$\beta p(\beta,h) = \lim_{\Lambda\to\infty}\frac{1}{\Lambda}\log\sum_{\underline{\sigma}} e^{-\beta H(\underline{\sigma})-h\sum_x\sigma_x}\,; \tag{5.10.11}$$

however the physical interpretations of f, p are of course different. To find the meaning of the above quantities in the Thermodynamics of a spin system one would have to go through the discussion of the orthodicity again, in the case of such systems. One would find, as it is easy to check, that $p(\beta,h)$ has the interpretation of *magnetic free energy* while $f(\beta,m)$ is a quantity that does not have a special name in the Thermodynamics of magnetic systems.

In the next chapter we shall consider some special cases: they are the simplest and they are quite remarkable as in some particular instances they are even amenable to more or less exact solution (*i.e.* calculation of the thermodynamic limit of various quantities, like for instance the free energy). The interest, as it will appear, of such models will be the wealth of information that they provide about the phenomena related to phase transitions.

One of the developments of the late 1960s and early 1970s is that natural "extensions" to lattice spin systems of the formalism discussed in §5.3-§5.7 arise in rather unexpected contexts, see Chap.IX.

Such "extended" lattice system models are spin systems more general than the model (5.10.4). For instance they allow the spins σ_x at each site x to be

an "arbitrary" finite set of symbols or "spin values" (rather than necessarily $\sigma_x = \pm 1$). The number of values of such "spins" divided by 2 is called the *total spin*: so that the case $\sigma = \pm 1$ is the "spin $\frac{1}{2}$" case.

It is convenient, for later reference purposes, to introduce here such extensions: they will be one-dimensional models without phase transitions at least in the cases that we shall later consider in Chap.IX; they can be extended also to higher dimension and, as such, they will appear in Chap.VII. The model energy H has the form

5.10.12
$$H = \sum_{n=1}^{\infty} \sum_{x_1 < \ldots < x_n} \varphi_{x_1,\ldots,x_n}(\sigma_{x_1}, \ldots, \sigma_{x_n}) \tag{5.10.12}$$

where $\varphi_{x_1,\ldots,x_n}(\sigma_1, \ldots, \sigma_n) \equiv \varphi_{x_1+a,\ldots,x_n+a}(\sigma_1, \ldots, \sigma_n)$ for all $a \in Z$ ("translation invariance"), and

5.10.13
$$\sum_{n=1}^{\infty} \sum_{0=x_1 < \ldots < x_n} w_n \max_{\sigma_1,\ldots,\sigma_n} |\varphi_{x_1,\ldots,x_n}(\sigma_1, \ldots, \sigma_n)| < \infty \tag{5.10.13}$$

for some weights $w_n \ge 1$. If $w_n = e^{\kappa |x_n - x_1|}$ one says that the model (5.10.12) is a *short-range Ising model with many body interactions.* The quantity:

5.10.14
$$\lambda(\underline{\sigma}) = \sum_{n=1}^{\infty} \sum_{0=x_1 < \ldots < x_n} \varphi_{x_1,\ldots,x_n}(\sigma_{x_1}, \ldots, \sigma_{x_n}) \tag{5.10.14}$$

will be called the "energy per site"; a few properties of λ should be noted. Namely λ is "Hölder continuous": *i.e.* if $\underline{\sigma}, \underline{\sigma}'$ are two spin configurations agreeing for $|i| \le k$ (*i.e.* $\sigma_i = \sigma'_i$ for all $|i| \le k$), and if the interaction has short range in the above sense, then for some $\kappa > 0$,

5.10.15
$$|\lambda(\underline{\sigma}) - \lambda(\underline{\sigma}')| < const\, e^{-\kappa k}\,. \tag{5.10.15}$$

This means that $\lambda(\underline{\sigma})$ depends "*exponentially little*" on the spins located far from the origin.

The partition function of the model with "open" boundary conditions will be simply

5.10.16
$$Z = \sum_{\underline{\sigma}} e^{-\beta H(\underline{\sigma})}\,. \tag{5.10.16}$$

More generally one can consider the model in the presence of "fixed spin boundary conditions". This means that for each configuration $\underline{\sigma}^\Lambda$ in the box $\Lambda = [-L, L]$ we consider the *biinfinite configuration* $\underline{\sigma} = (\underline{\sigma}^L, \underline{\sigma}^\Lambda, \underline{\sigma}^R)$ obtained by putting $\underline{\sigma}^\Lambda$ on the lattice and then continuing it outside Λ with $\underline{\sigma}^L$ to the left and $\underline{\sigma}^R$ to the right, where $\underline{\sigma}^L = (\ldots, \sigma^L_{-1}, \sigma^L_0)$ and $\sigma^R = (\sigma^R_0, \sigma^R_1, \ldots)$. The probability $\mu(\underline{\sigma}^\Lambda)$ of a configuration in the model "with boundary conditions $\underline{\sigma}_L, \underline{\sigma}_R$" will be:

5.10.17
$$\mu(\underline{\sigma}^\Lambda) = \frac{e^{-H(\underline{\sigma}^L \underline{\sigma}^\Lambda \underline{\sigma}^R)}}{\sum_{\underline{\sigma}'^\Lambda} e^{-H(\underline{\sigma}^L \underline{\sigma}'^\Lambda \underline{\sigma}^R)}} \tag{5.10.17}$$

where, if ϑ denotes the shift operation on the bilateral sequences and we set $\underline{\sigma}^L\underline{\sigma}^\Lambda\underline{\sigma}^R=\underline{\sigma}$, the energy H is

5.10.18
$$H(\underline{\sigma}^L\underline{\sigma}^\Lambda\underline{\sigma}^R)=H(\underline{\sigma})=\sum_{k=-L}^{L}\lambda(\vartheta^k\underline{\sigma}) \tag{5.10.18}$$

so that one can remark that the exponential is simply written in terms of the energy per site $\lambda(\underline{\sigma})$; (5.10.12) can also be expressed in a similar way.

In fact let $\underline{\sigma}=\underline{\sigma}^{o,\Lambda}$ be the biinfinite configuration obtained by extending $\underline{\sigma}^\Lambda$ periodically outside the region Λ. Then the energy of interaction between the spins in Λ and between them and the ones outside Λ, is:

5.10.19
$$H(\underline{\sigma}^\Lambda)=\sum_{k=-L}^{L}\lambda(\vartheta^k\underline{\sigma})+corrections \tag{5.10.19}$$

where the "corrections" depend "only" on the spins near the boundary points $\pm L$ and outside the interval $\Lambda=[-L,L]$, in the sense that by varying the spin at a site at distance ℓ from the boundary the correction changes by a quantity proportional to $e^{-\kappa\ell}$ if κ is the exponent in the weight w_n introduced above, see (5.10.13).

A further extension is obtained by considering a matrix T whose entries $T_{\sigma\sigma'}$ are labeled by the spin values and are supposed to be $T_{\sigma\sigma'}=0,1$, and by restricting the family of spin configurations $\underline{\sigma}$ to the T*-compatible configurations*: they are defined to be those that satisfy $T_{\sigma_i,\sigma_{i+1}}=1$, where T is a matrix with entries 0 or 1, called the *compatibility matrix*. One also calls such models "*hard core spin systems*" and T describes the hard core structure. Such systems are also called *hard core lattice systems*.

If the matrix T is such that $T^n_{\sigma,\sigma'}>0$, for all σ,σ' and for n large enough, one says that the hard core is *mixing*. For such cases all the above formulae and definitions (5.10.12)-(5.10.19) extend unchanged provided only compatible spin configurations are considered.

Appendix 5.A1: Absence of Phase Transition in non Nearest Neighbor One-Dimensional Systems

The method discussed in §5.8 for hard core nearest neighbour one-dimensional models is called the *transfer matrix* method. We extend it here to the more general case of finite range, but not nearest neighbor. The theory is very similar to that in §5.8.

Let $\varphi(r)\equiv 0$ for $r>(n-1)r_0$ for some integer n. Assume for simplicity that N is a multiple of n. Let $\underline{q}=(q_1,\ldots,q_n)$ with $q_1<q_2<\ldots<q_n$; and let us define

5.A1.1
$$a(\underline{q})=a(q_1,\ldots,q_n)=\frac{1}{2}\sum_{i<j=1}^{n}\varphi(q_j-q_j) \tag{5.A1.1}$$

$$b(\underline{q}\,|\,\underline{q}') = b(q_1,\ldots,q_n\,|\,q'_1,\ldots,q'_n) = \sum_{i,j=1}^{n}\varphi(q_i-q'_j)$$
$$c(\underline{q}\,|\,\underline{q}') = a(\underline{q}) + b(\underline{q}\,|\,\underline{q}') + a(\underline{q}')\,.$$

Then the energy of a particle configuration is

5.A1.2
$$a(q_1,\ldots,q_n)+a(q_{n((N/n)-1)+1},\ldots,q_N)+ \\ +\sum_{k=0}^{N/n-2} c(q_{nk+1}\ldots q_{nk+n}\,|\,q_{n(k+1)+1},\ldots,q_{n(k+2)}) \tag{5.A1.2}$$

so that one easily finds

5.A1.3
$$J_N(\beta,p)=\langle\, a\,|\,T^{\frac{N}{n}-1}\,|a\,\rangle \tag{5.A1.3}$$

where T is the operator acting on the space of the functions of n coordinates $\underline{q}=(q_1,q_2,\ldots,q_n)$ with $q_{j+1}-q_j>r_0$:

5.A1.4
$$Tf(\underline{q})=\int_0^\infty e^{-\beta c(\underline{q}\,|\,q_n+\underline{q}')}e^{-\frac12(q_n+q'_1)\beta p}f(\underline{q}')d\underline{q}' \tag{5.A1.4}$$

and the vector $|a\,\rangle$ is the function $e^{-\frac12\beta p q_n-\beta a(\underline{q})}$.

Since the operator T is a Hilbert-Schmidt operator on the space $L_2(d\underline{q})$ (*i.e.* $\int T(\underline{q},\underline{q}')^2 d\underline{q}d\underline{q}'<+\infty$), and since its kernel is >0 it "immediately" follows (*i.e.* it follows from well known results on the theory of operators, or better of matrices, like the *Perron-Frobenius theorem*, see p. 136 in [Ru69]) that the largest eigenvalue $t(\beta,p)$ of T is isolated and simple and therefore it is analytic as a function of β,p, since T itself is analytic in such variables.

Therefore $\lambda(\beta,p)=\frac1n t(\beta,p)$ is analytic in β,p for $\beta,p>0$, and convex in such variables (see (5.8.2) showing that J_N is a "linear combination" of functions depending on β as $e^{\beta c}$, hence $\log J_N(\beta,p)$, is convex in β) and we can repeat the argument above to see that the equation of state gives p as an analytic function of β,v.

The further extension to systems with a potential with infinite range but *satisfying* (5.8.4) is also possible and it was the main purpose of van Hove's theorem, [VH50]. The condition (5.8.4) comes in to insure that $b(\underline{q}\,|\,\underline{q}')$ is *uniformly bounded*: this quantity represents the interaction between a configurartion $\underline{q}$ situated *to the left* of another configuration $\underline{q}'$, hence it is uniformly bounded if (5.8.4) holds.

.

Chapter VI

Coexistence of Phases

.

§6.1. The Ising Model. Inequivalence of Canonical and Grand Canonical Ensembles[1]

The Ising model[2] plays a very special role in statistical mechanics and generates the simplest nontrivial example of a system undergoing phase transitions.

Its analysis has provided us with deep insights into the general nature of phase transitions, which are certainly better understood nowadays, after the publication of the hundreds of papers which followed the pioneering work of Ising, Peierls, Onsager, Kaufman and Yang, [Pe36], [On44], [Ka49], [KO49], [Ya52].

The main reason why so much attention has been given to this very special model lies in its simplicity and, in spite of it, in the fact that it first gave firm and quantitative indications that a microscopic short-range interaction can produce phase transitions which, furthermore, deeply differ in character from the classical van der Waals' (or Curie-Weiss' or mean field) type of transitions, see §5.1 and §5.2.

It should also be mentioned that the two-dimensional Ising model in zero external field is exactly solvable (see §7.4);[3] this fact has been very often used to check of the validity of numerical approximations designed for applications to more complicated models, see the review [Fi64], pp. 677-702.

Last but not least, we mention that the Ising model has given rise to a number of interesting developments and reinterpretations of old results in the theory of Markov chains, [Do68],[Sp71], information theory, [Ru69], [Or74], [RM75], random walks, [Gr67],[Fi67b],[GH64], [La85], to quote a few remarkable works, and therefore constitutes a notable example of a subject which has simultaneously been the object of advanced research in Physics, Mathematics and Mathematical Physics.

In the rest of this chapter we give a description, certainly not exhaustive,

---

1 This chapter is mostly taken from the paper *Instabilities and phase transitions in the Ising model*, *La Rivista del Nuovo Cimento*, **2**, 133–169, 1972.

2 For a history of the Ising model see [Br69].

3 The original solution for the free energy of the Ising model in two dimensions can be found in [On44]. It was preceded by the proofs of existence of Peierls, [Pe36], and van der Waerden, [VW41], and by the exact location of the critical temperature by Kramers and Wannier, [KW41].
The spontaneous magnetization was found by Onsager, [KO49], but the details were never published; it was subsequently rediscovered by Yang, [Ya52]. A modern derivation of the solution is found in the review article by Schultz, Mattis and Lieb, [SML64]: the latter is reproduced in §7.4. Another interesting older review article is the paper [NM53]. A combinatorial solution has been found by Kac and Ward and can be found in [LL67],p. 538. Some aspects of this derivation were later clarified: and it has been discussed again in several papers, see [Be69]. Another approach to the solution (the Kasteleyn's *pfaffian method*) can be found in [Ka61].

of the model and of some selected results. They illustrate properties which throw some light on the general nature of the phenomenon of phase transitions, mostly far from the critical point, and which, hopefully, should not be a peculiarity of the simplicity of the model.

There exist some very good accounts of the theoretical arguments leading to the consideration of the Ising model in the context of physical problems, [Fi67a], [Ma65]; here we shall completely skip this aspect of the matter.[4]

§6.2. The Model. Grand Canonical and Canonical Ensembles. Their Inequivalence

We consider a d-dimensional ($d = 1, 2, 3$) square lattice Z^d and a finite square $\Lambda \subset Z^d$ centered around the origin, containing $|\Lambda| = L^d$ lattice sites.

On each site $x \in \Lambda$ is located a classical "*spin*" $\sigma_x = \pm 1$. The "*configurations*" of our system will, therefore, consist of a set $\underline{\sigma} = (\sigma_{x_1}, \ldots, \sigma_{x_{|\Lambda|}})$ of $|\Lambda|$ numbers $\sigma_x = \pm 1$; the number of these configurations is $2^{|\Lambda|}$. The ensemble of the configurations will be denoted $\mathcal{U}(\Lambda)$.

To each spin configuration a certain *energy* is assigned, see §5.10:

6.2.1
$$H_\Lambda(\underline{\sigma}) = -J \sum_{<i,j>} \sigma_{x_i}\sigma_{x_j} - h \sum_i \sigma_{x_i} + \mathcal{B}_\Lambda(\underline{\sigma}) \tag{6.2.1}$$

where $\sum_{<i,j>}$ means that the sum is over pairs (x_i, x_j) of neighboring points, h is an "*external magnetic field*" and $\mathcal{B}_\Lambda(\underline{\sigma})$ describes the interaction of the spins in the box Λ with the "rest of the world". This could be the contribution to the energy that comes from the *fixed spins* boundary conditions that we considered in §5.5.[5]

For simplicity we shall treat only the case $J > 0$.

Of course $\mathcal{B}_\Lambda(\underline{\sigma})$ in (6.2.1) can be rather arbitrary and, in fact, depends on the particular physical problem under investigation. It is subject, however, to one constraint of physical nature: in case we were interested in letting $\Lambda \to \infty$, we should impose the condition:

6.2.2
$$\lim_{\Lambda\to\infty} \frac{\max_{\underline{\sigma}} |\mathcal{B}_\Lambda(\underline{\sigma})|}{|\Lambda|} = 0 \tag{6.2.2}$$

i.e. we want the condition that the energy due to $\mathcal{B}_\Lambda(\underline{\sigma})$ should not be of the same order as the volume of the box; *furthermore* $\mathcal{B}_\Lambda$ should depend mostly on the σ_x with x near the boundary; *e.g.* $\mathcal{B}_\Lambda(\underline{\sigma}) = c\sigma_0$ satisfies (6.2.2) but it should also be excluded.[6] In other words $\mathcal{B}_\Lambda$ should be a "surface term".

---

[4] In some cases the Ising model is a good phenomenological model for antiferromagnetic materials: this is the case of $MnCl_2 \cdot 4H_2O$, see [FS62], [Fi67].

[5] This term is usually omitted and in some sense its importance has only recently been recognized after the work of Dobrushin, Lanford and Ruelle, see [Do68], [LR69]. In this chapter the main purpose is to emphasize the role of this term in the theory of phase transitions.

[6] A precise condition could be that for any fixed set D, $\max |\mathcal{B}_\Lambda(\sigma) - \mathcal{B}_\Lambda(\underline{\sigma}')| \xrightarrow[\Lambda\to\infty]{} 0$

The laws of statistical mechanics provide a relationship between the microscopic Hamiltonian (6.2.1) and the macroscopic quantities appearing in the thermodynamical theory of the system. The free energy per unit volume is given by

6.2.3
$$f_\Lambda(\beta,h)=\frac{\beta^{-1}}{|\Lambda|}\log Z(\beta,h,\Lambda,\mathcal{B}) \tag{6.2.3}$$

where $\beta=T^{-1}$ is the inverse temperature and

6.2.4
$$Z(\beta,h,\Lambda,\mathcal{B})=\sum_{\underline{\sigma}\in\mathcal{U}(\Lambda)} e^{-\beta H_\Lambda(\underline{\sigma})} \tag{6.2.4}$$

is the grand canonical partition function. Furthermore the probability of finding the system in a configuration $\underline{\sigma}$ of the grand canonical ensemble $\mathcal{U}(\Lambda)$ is given by the Boltzmann factor:

6.2.5
$$\frac{e^{-\beta H_\Lambda(\underline{\sigma})}}{Z(\beta,h,\Lambda,\mathcal{B})},\qquad \underline{\sigma}\in\mathcal{U}(\Lambda)\,. \tag{6.2.5}$$

The grandcanonical ensemble formalism based on (6.2.3),(6.2.5) corresponds to the physical situation in which there are no constraints on the system. If one could, by some experimental arrangement, regard for example the total magnetization $M(\underline{\sigma})=\sum_{x\in\Lambda}\sigma_x$ as fixed: $M(\underline{\sigma})=M=m|\Lambda|$, then the expression (6.2.3) for the free energy would no longer be appropriate.

One should rather consider the canonical ensemble, *i.e.* the set of the allowed configurations would be the set $\mathcal{U}(\Lambda,m)\subset\mathcal{U}(\Lambda)$ consisting of all the $\underline{\sigma}\in\mathcal{U}(\Lambda)$ such that $\sum_{x\in\Lambda}\sigma_x=m|\Lambda|$, ($|m|<1$), and the Thermodynamics would be described by the function

6.2.6
$$g_\Lambda(\beta,h,m)=\frac{\beta^{-1}}{|\Lambda|}\log Z(\beta,h,\Lambda,\mathcal{B},m) \tag{6.2.6}$$

where

6.2.7
$$Z(\beta,h,\Lambda,\mathcal{B},m)=\sum_{\underline{\sigma}\in\mathcal{U}(\Lambda,m)} e^{-\beta H_\Lambda(\underline{\sigma})} \tag{6.2.7}$$

and the free energy would be $\tilde f_\Lambda(\beta,h)$:

6.2.8
$$\tilde f_\Lambda(\beta,h)=h\,m(h)+g_\Lambda(\beta,0,m(h))\,. \tag{6.2.8}$$

where $m(h)$ is the solution of the equation:[7]

6.2.9
$$h=-\frac{\partial g_\Lambda(\beta,0,m)}{\partial m} \tag{6.2.9}$$

---

if the maximum is over all pairs of spin configurations $\underline{\sigma}$ and $\underline{\sigma}'$ that differ only on D, *i.e.* such that $\sigma_x=\sigma'_x$ for $x\notin D$.

[7] Here we have not been precise about the problem of what $\partial/\partial m$ means, since $g_\Lambda(\beta,m)$ is only defined for certain rational values of m (whose number is finite). One could, for instance, extend $g_\Lambda(\beta,m)$ to all m's by considering instead of $g_\Lambda(\beta,m)$ its convex envelope (or also one could prefer to consider the g_Λ obtained by linear interpolation from (6.2.6)). This is not very satisfactory but it should not be very important for large systems, as discussed in Chap.IV.

There is no reason for having $\tilde{f}_\Lambda = f_\Lambda$ since they correspond to different physical problems; it is only when, *in some sense*, the fluctuations become negligible (*i.e.* in the limit $\Lambda \to \infty$) that one can expect the identity between $\tilde{f}$ and f.

Of course in general the difference between $\tilde{f}_\Lambda$ and f_Λ should vanish as $|\Lambda|^{-1}$ times $O(|\Lambda|^{(d-1)/d})$ (this means $O(\log|\Lambda|)$ for $d=1$); but, as we shall see on many occasions, the situation is not so simple for other quantities such as the correlation functions or the average magnetization.

As discussed in §5.9 the inequivalence, for finite volume, of the predictions of the canonical and grand canonical ensembles should not be interpreted as meaning that statistical mechanics is only approximate when applied to finite systems; it simply means that in dealing with finite systems attention must be paid to the boundary conditions as a manifestation of the peculiarities of the actual physical situation from which the problem under consideration arises. We conclude by remarking that in the canonical ensemble the probability of a spin configuration will be given by an expression similar to (6.2.5):

6.2.10
$$\frac{e^{-\beta H_\Lambda(\underline{\sigma})}}{Z(\beta,h,\Lambda,\mathcal{B},m)}, \qquad \underline{\sigma} \in \mathcal{U}(\Lambda,m)\,. \tag{6.2.10}$$

§6.3. Boundary Conditions. Equilibrium States

Formulae (6.2.5), or (6.2.10), provide a complete statistical description of the properties of the system. An alternative and often more convenient, equally complete, description is provided by the so-called correlation functions:

6.3.1
$$\langle \sigma_{x_1}\sigma_{x_2}\ldots\sigma_{x_n}\rangle_{\Lambda,\mathcal{B}_\Lambda} = \frac{\sum_{\underline{\sigma}} \sigma_{x_1}\sigma_{x_2}\ldots\sigma_{x_n}\, e^{-\beta H_\Lambda(\underline{\sigma})}}{\sum_{\underline{\sigma}} e^{-\beta H_\Lambda(\underline{\sigma})}} \tag{6.3.1}$$

where $\sum_{\underline{\sigma}}$ is extended to the appropriate statistical ensemble. For instance the average magnetization in the grand canonical ensemble $\mathcal{U}(\Lambda)$ is

6.3.2
$$m_\Lambda(\beta,h) = \frac{\partial f_\Lambda(\beta,h)}{\partial h} = \frac{\sum_{x\in\Lambda}\langle\sigma_x\rangle_{\Lambda,\mathcal{B}}}{|\Lambda|} \tag{6.3.2}$$

We shall refer to the family of correlation functions (6.3.1) (regarded as a whole) as the "equilibrium state of the system in the box Λ".

We call an *equilibrium state*, see §5.5, of the infinite system any family $\{\langle\sigma_{x_1}\ldots\sigma_{x_n}\rangle\}$ of functions such that, for a suitable choice of the $\mathcal{B}_\Lambda(\underline{\sigma})$:

6.3.3
$$\langle\sigma_{x_1}\ldots\sigma_{x_n}\rangle = \lim_{\Lambda\to\infty}\langle\sigma_{x_1}\ldots\sigma_{x_n}\rangle_{\Lambda,\mathcal{B}_\Lambda} \tag{6.3.3}$$

for all $n \geq 1$ and all $x_1, x_2, \ldots, x_n \in Z^d$, *simultaneously.*[8]

An equilibrium state for an infinite system will simply be called an equilibrium state: it is specified by a suitable choice of a sequence $\{\mathcal{B}_\Lambda(\underline{\sigma})\}$ of boundary conditions satisfying the requirement (6.2.2).

Let us list a number of remarkable boundary conditions:

(1) *Open boundary condition* (also called "perfect-wall" boundary conditions). This name will be given to the case

6.3.4 $$\mathcal{B}_\Lambda(\underline{\sigma}) \equiv 0 \qquad \text{for all } \underline{\sigma} \in \mathcal{U}(\Lambda) \tag{6.3.4}$$

(2) *Periodic boundary conditions.* This corresponds to allowing spins on opposite faces of the box Λ to interact through a coupling $-J$ (*i.e.* as the bulk spins). Clearly this can be obtained by a suitable choice of $\mathcal{B}_\Lambda(\underline{\sigma})$; we shall refer to this choice as "periodic boundary conditions".

(3) $(\underline{\varepsilon})$-*boundary conditions.* Let $(\xi_1, \xi_2, \ldots)$ be the $2d|\Lambda|^{(d-1)/d}$ lattice points adjacent to the boundary of Λ. Let $\underline{\varepsilon} = (\varepsilon_{\xi_1}, \varepsilon_{\xi_2}, \ldots)$, $\varepsilon_{\xi_i} = \pm 1$, be fixed. We shall call (ε)-boundary condition the choice

6.3.5 $$\mathcal{B}_\Lambda(\underline{\sigma}) = -J \sum_{x_i \in \partial\Lambda} \sigma_{x_j} \varepsilon_{\xi_j} \tag{6.3.5}$$

where (x_i, ξ_i) are nearest neighbors.

The physical meaning of this boundary condition is clear: we imagine that the sites neighboring the boundary $\partial\Lambda$ of Λ are occupied by a spin configuration $\underline{\varepsilon}$ and that the latter spins interact with the spins $\underline{\sigma}$ through the same coupling constant of the bulk spins.

The cases $\underline{\varepsilon} = (+1, +1, \ldots, +1)$ or $\underline{\varepsilon} = (-1, -1, \ldots, -1)$ will be, respectively, referred to as the $(+)$-*boundary condition* or the $(-)$-*boundary condition.*

(4) In the two-dimensional case we shall be interested in another boundary condition. Suppose that the spins on the opposite *vertical* sides of Λ are allowed to interact through a coupling $-J$ (*i.e.* we impose periodic boundary conditions along the rows of Λ *only*); and suppose that a set $\underline{\varepsilon}_u$ of fixed spins is located on the lattice sites adjacent to the upper base of Λ and, similarly, a set $\underline{\varepsilon}_l$ of fixed spins is adjacent to the lower base of Λ. The spins $\underline{\varepsilon}_u, \underline{\varepsilon}_l$ are allowed to interact with the nearest spins in Λ with a coupling $-J$. We shall

---

[8] This definition is essentially in [LR69] where the equivalence of the above definition with a number of other possible definitions is shown. For instance the definition in question is equivalent to that based on the requirement that the correlation functions should be a solution of the equations for the correlation functions that can be derived for lattice gases or magnetic spin systems in analogy to those we discussed for the gases in §5.8. It is also equivalent to the other definitions of equilibrium state in terms of tangent planes (*i.e.* functional derivatives of a suitable functional: see [Ru69], p. 184, [Ga81]).

naturally refer to this choice of $\mathcal{B}_\Lambda(\underline{\sigma})$ as the $(\underline{\varepsilon}_u, \underline{\varepsilon}_l)$-*cylindrical boundary condition*.

The particular cases

6.3.6
$$\begin{aligned} \underline{\varepsilon}_u =&(+1,+1,\ldots,+1), \qquad \underline{\varepsilon}_l = (+1,+1,\ldots,+1)\\ \underline{\varepsilon}_u =&(+1,+1,\ldots,+1), \qquad \underline{\varepsilon}_l = (-1,-1,\ldots,-1)\end{aligned} \tag{6.3.6}$$

will be referred to, respectively, as $(+,+)$-*cylindrical boundary* condition or $(+,-)$-*cylindrical boundary condition*.

§6.4. The Ising Model in One and Two dimensions and zero field

To acquire some familiarity with the model we examine some of the simplest cases. Consider the one-dimensional Ising chain with periodic boundary conditions. Labeling points of Λ as $1, 2, \ldots, L$, the zero field Hamiltonian is

6.4.1
$$H_\Lambda(\underline{\sigma}) = -J \sum_{i=1}^{L} \sigma_i \sigma_{i+1}, \qquad \sigma_{L+1} \equiv \sigma_1 \tag{6.4.1}$$

(clearly $\mathcal{B}_\Lambda(\underline{\sigma}) = -J\sigma_L\sigma_1$). The grand canonical partition function can be written:

6.4.2
$$Z_\Lambda(\beta) = \sum_{\underline{\sigma}} e^{\beta J \sum_{i=1}^{L} \sigma_i\sigma_{i+1}} = \sum_{\underline{\sigma}} \prod_{i=1}^{L} e^{\beta J \sigma_i \sigma_{i+1}}\,. \tag{6.4.2}$$

Noting that $(\sigma_i\sigma_{i+1})^2 \equiv 1$ and therefore

6.4.3
$$e^{\beta J \sigma_i\sigma_{i+1}} \equiv \cosh\beta J + \sigma_i\sigma_{i+1} \sinh\beta J \tag{6.4.3}$$

(6.4.2) can be rewritten as

6.4.4
$$Z_\Lambda(\beta) = (\cosh\beta J)^L \sum_{\underline{\sigma}} \prod_{i=1}^{L} (1 + \tanh\beta J\, \sigma_i\sigma_{i+1})\,. \tag{6.4.4}$$

If one develops the product in (6.4.4) one gets a sum of terms of the form

6.4.5
$$(\tanh\beta J)^k \sigma_{i_1}\sigma_{i_1+1}\sigma_{i_2}\sigma_{i_2+1}\ldots\sigma_{i_k}\sigma_{i_k+1}\,. \tag{6.4.5}$$

It is clear that, unless $k = 0$ or $k = L$, each of the terms (6.4.5) contains at least an index i_j which appears only once. Therefore, after performing the sum over the $\underline{\sigma}$'s, all terms (6.4.5) give a vanishing contribution to $Z_\Lambda(\beta)$ except the two with $k = 0$ and $k = L$ which are, respectively, 1 and $(\tanh\beta J)^L \cdot\, \sigma_1\sigma_2\sigma_2\ldots\sigma_{L-1}\sigma_L\sigma_L\sigma_1 \equiv (\tanh\beta J)^L$. This implies

6.4.6
$$Z_\Lambda(\beta) = (\cosh\beta L)^L 2^L\,(1 + (\tanh\beta J)^L) \tag{6.4.6}$$

Hence:[9]

6.4.7 $$\beta f_\Lambda(\beta) = \log(2\cosh\beta J) + \frac{1}{L}\log(1 + (\tanh\beta J)^L)\,. \qquad (6.4.7)$$

It has to be remarked that $\beta f_\Lambda(\beta)$, as well as $\beta f(\beta) = \lim_{L\to\infty}\beta f_\Lambda(\beta) = \log 2\cosh\beta J$, are analytic in β; this fact is usually referred to as the "*absence of phase transitions*" in the one-dimensional Ising model. The reader can check, by using the above method, that the partition function in the grand canonical ensemble and zero field but open boundary conditions (see §6.3) is slightly different from (6.4.6) and, precisely, is equal to $(\cosh\beta J)^L 2^L$.

Consider now the two-dimensional Ising model in a zero field and with open boundary conditions:

6.4.8 $$H_\Lambda(\underline{\sigma}) = -J\sum_{i=1}^{L}\sum_{j=1}^{L-1}\sigma_{i\,j}\sigma_{i\,j+1} - J\sum_{i=1}^{L-1}\sum_{j=1}^{L}\sigma_{i\,j}\sigma_{i+1\,j}\,. \qquad (6.4.8)$$

A better form for $H_\Lambda(\underline{\sigma})$ is the following:

6.4.9 $$H_\Lambda\underline{\sigma} = -J\sum_b \tilde{\sigma}_b \qquad (6.4.9)$$

where $\sum_b$ denotes the sum over the bonds, *i.e.* over the segments $b = [(i,j),(i,j\ +1)]$ or $b = [(i,j),(i+1,j)]$, and $\tilde{\sigma}_b$ is the product of the two spins at the extremes of b (*e.g.* if $b = [(i,j),(i+1,j)]$ then $\tilde{\sigma}_b = \sigma_{i\,j}\sigma_{i+1,j}$).

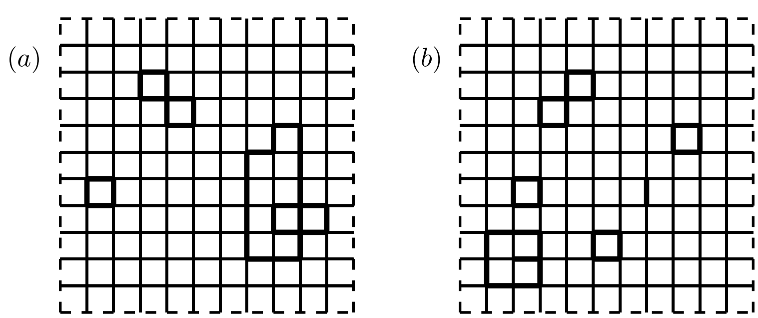
The partition function can be written, as in the one-dimensional case, as

6.4.10 $$Z_\Lambda(\beta) = (\cosh\beta J)^{2L(L-1)}\sum_{\underline{\sigma}}\prod_b(1 + (\tanh\beta J)\tilde{\sigma}_b) \qquad (6.4.10)$$

Developing the product we are led to a sum of terms of the type:

6.4.11 $$(\tanh\beta J)^k\tilde{\sigma}_{b_1}\tilde{\sigma}_{b_2}\ldots\tilde{\sigma}_{b_k} \qquad (6.4.11)$$

and we can conveniently describe this term through the geometric set of lines $b_1, b_2, \ldots, b_k$. After the $\sum_{\underline{\sigma}}$ is taken, many terms of the form (6.4.1) give a vanishing contribution. The ones that give a nonvanishing contribution are those in which the vertices of the geometric figure $b_1 \cup b_2 \cup \ldots \cup b_k$ belong to an even number of b_j's (two or four).



[9] The solution can also be found for instance in [NM53].

Fig. 6.4.1: The dashed line is the boundary of Λ.

These terms are the ones such that $\tilde\sigma_{b_1}\cdot\tilde\sigma_{b_2}\ldots\tilde\sigma_{b_k}\equiv 1$. In Fig. 6.4.1a we give a typical nonvanishing term and in Fig. 6.4.1b an example of a vanishing term ($k=30$).

We shall, in the following, consider the geometric figures built with k segments $b_1,\ldots,b_k$ such that $\tilde\sigma_{b_1}\cdot\tilde\sigma_{b_2}\ldots\tilde\sigma_{b_k}\equiv 1$ and call it a *k-sided multi-polygon* on the box Λ (needless to say, all the $b_1,\ldots,b_k$ are pairwise different). Let $P_k(\Lambda)$ be the number of such polygons.

The partition function is now easily written as[10]:

6.4.12
$$Z_\Lambda(\beta)=(\cosh\beta J)^{2L(L-1)}2^{L^2}\sum_{k\ge 0}P_k(\Lambda)\,(\tanh\beta J)^k\,. \qquad (6.4.12)$$

§6.5. Phase Transitions. Definitions

We have already seen, in the preceding section, that the one dimensional Ising model has no phase transitions in zero field, since both $f_\Lambda(\beta)$ and $f(\beta)$ are analytic in β.

We recall briefly in the concrete context of the Ising model the general considerations of Chap.V about the definition of phase transition as a phenomenon of macroscopic instability: slight changes of external conditions should imply dramatic changes of some macroscopic variables; it is hard to imagine how in such a situation thermodynamic functions, which we have seen to be boundary-condition independent, like the free energy, the pressure, *etc*, could be analytic functions of the parameters in terms of which they are expressed (say, temperature, chemical potential or magnetic field, *etc*).

For this reason an analytic singularity in the thermodynamic functions is usually thought of as a "symptom" of a phase transition and on this idea it would be possible to base a definition and a theory of the phenomenon of phase transitions.

Here, however, we will not base the investigation of the nature of phase transitions in the Ising model on the search for singularities of the thermodynamic functions; we shall rather adopt and make more precise the other, perhaps more immediate and intuitive, approach based on the detection of "macroscopic instabilities", introduced in Chap.V.

This way of proceeding is more convenient for the simple reason that a number of very clear and rather deep results have been obtained along these lines. But it should be understood that this second approach does not "brilliantly" avoid the difficulties of the first. It is simply an approach to the theory of phase transitions which, so far, has asked for and provided a less refined description of the phenomena of interest, as compared to the

[10] The expansion can be used as a starting point for the combinatorial solution mentioned above, see [LL67].

description which would be expected from the analysis of the singularities of appropriate analytic functions (an analysis still in a very primitive stage and whose problems are often not well formulated even in the simplest cases).[11] For this reason it provides a wealth of remarkable properties of the phase transition phenomenon.

Let us now discuss in a more precise way the concept of macroscopic instability. Consider the Ising model and define the condition that a phase transition takes place at the values (β, h) of the thermodynamic parameters if the system is unstable with respect to boundary perturbations; *i.e.* if there are at least two sequences $\mathcal{B}_\Lambda(\underline{\sigma})$ and $\mathcal{B}'_\Lambda(\underline{\sigma})$ of boundary terms (see (6.2.1), (6.2.2)) such that (say, in the grand canonical ensemble)

6.5.1 $$\lim_{\Lambda\to\infty} \langle \sigma_{x_1} \ldots \sigma_{x_n} \rangle_{\Lambda,\mathcal{B}_\Lambda} \neq \lim_{\Lambda\to\infty} \langle \sigma_{x_1} \ldots \sigma_{x_n} \rangle_{\Lambda,\mathcal{B}'_\Lambda} \tag{6.5.1}$$

for a suitable choice of $x_1, x_2, \ldots, x_n, n$.

We first clarify why we say that, if (6.5.1) holds, we have a macroscopic instability. We remark that a change in boundary conditions does not change the extensive properties of the system such as the free energy. In fact, from definition (6.2.4):

6.5.2 $$\frac{Z(\beta,h,\Lambda,\mathcal{B}_\Lambda)}{Z(\beta,h,\Lambda,\mathcal{B}'_\Lambda)} \le e^{\max_{\underline{\sigma}\in\mathcal{U}(\Lambda)} |\mathcal{B}_\Lambda(\underline{\sigma})|+|\mathcal{B}'_\Lambda(\underline{\sigma})|} \tag{6.5.2}$$

and therefore (6.2.2) implies

6.5.3 $$\lim_{\Lambda\to\infty} \frac{1}{|\Lambda|} \log Z(\beta,h,\Lambda,\mathcal{B}_\Lambda) \equiv \lim_{\Lambda\to\infty} \frac{1}{|\Lambda|} Z(\beta,h,\Lambda,\mathcal{B}'_\Lambda) \tag{6.5.3}$$

On the other hand, if (6.5.1) is true, intensive quantities like the correlation functions are sensitive to the boundary conditions; for instance if

6.5.4 $$\lim_{\Lambda\to\infty} \langle \sigma_{x_1} \rangle_{\Lambda,\mathcal{B}_\Lambda} \neq \lim_{\Lambda\to\infty} \langle \sigma_{x_1} \rangle_{\Lambda,\mathcal{B}'_\Lambda} . \tag{6.5.4}$$

we realize that the local magnetization changes as a consequence of a change in boundary condition *even* if the boundary is very remote.

Of course once provided with a "definition" of what a phase transition is, one has not gone very far. The real question is whether the definition reflects what is physically expected; this implies, in particular, that one should at least be able to prove the existence of a phase transition, in the above

[11] Of course we do not attach a deep physical meaning to the difference between these two approaches. Clearly they should be equivalent if one pretended to extract all possible information from them. What is really important is that the first questions raised by both approaches are very interesting and relevant from a physical point of view. One of the goals of the analytic theory of phase transitions is to understand the nature of the singularity at the critical point and at the "breaks" of the isotherms. A lot of interest has been devoted to this point and a number of enlightening phenomenological results are available. However the number of complete results on the matter is rather limited. An idea of the type of problems that are of interest can be obtained by reading the papers [Ka68] or the more detailed paper [Fi67].

sense, in cases in which one expects a transition. Hopefully the definition and its physical interpretation should allow one to do more: for instance to provide the tools for a closer description of typical phenomena (like the phase separation).

Here we end this somewhat philosophical but necessary discussion and in the coming sections we shall describe in the concrete example of the Ising model, some of the results that have been obtained since the early 1960s, when the above point of view was starting to be developed, quite independently, by several people.

§**6.6. Geometric Description of the Spin Configurations**

Here we introduce a new description of the spin configurations, which we shall use to derive in a very elegant way the exact value of the critical temperature in the two-dimensional Ising model. In the following sections the geometric representation, introduced below, will be widely used.

Consider an Ising model with boundary conditions of the type (6.3.5) (($\underline{\varepsilon}$)-boundary conditions) or with periodic boundary conditions (see §6.3).

Given a configuration $\underline{\sigma} \in \mathcal{U}(\Lambda)$ we draw a unit segment perpendicular to the center of each bond b having opposite spins at its extremes (in three dimensions we draw a unit square surface element perpendicular to b). A two-dimensional example of this construction is provided by Fig. 6.6.1 (where a very special ($\underline{\varepsilon}$)-boundary condition is considered).

The set of segments can be grouped into lines (or surfaces in three dimensions) which separate regions where the spins are positive from regions where they are negative.

It is clear that some of the lines (or surfaces, if $d = 3$) are "closed polygons" ("closed polyhedra", respectively) while others are not closed. It is perhaps worth stressing that our polygons are not really such in a geometrical sense, since they are not necessarily "*self-avoiding*" (see Fig. 6.6.1): however they are such that they can intersect themselves only on vertices (and not on sides). From a geometrical point of view a family of disjoint polygons (in the above sense and in two dimensions) is the same thing as a multi-polygon in the sense discussed in §6.4, Fig. 6.4.1.

In two dimensions instead of saying that a polygon is "closed" we could equivalently say that its vertices belong to either two or four sides.

We note that the (+)-boundary conditions, the (−)-boundary conditions and the periodic boundary conditions are such that the lines (surfaces) associated with spin configurations are *all* closed polygons (polyhedra). In the periodic case some polygons might wind up around the two holes of the torus.

In the two-dimensional case and if the boundary conditions are the $(+,+)$-cylindrical or the $(+,-)$-cylindrical ones (see §6.3) a geometric construction of the above type can still be performed and, also in this case, the lines are closed polygons (some of which may "wind around" the cylinder Λ).

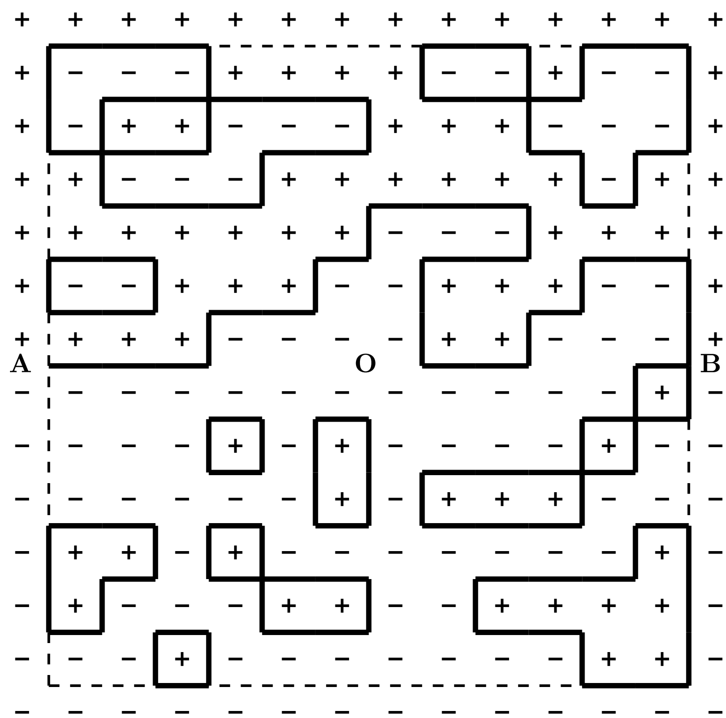


Fig. 6.6.1: The dashed line is the boundary of Λ; the outer spins are those fixed by the boundary condition. The points A, B are points where an open line ends.

For a fixed boundary condition let $(\gamma_1, \gamma_2, \ldots, \gamma_h, \lambda_1, \ldots, \lambda_k)$ be the disjoint components of the set of lines (surfaces) associated by the above construction with a spin configuration $\underline{\sigma} \in \mathcal{U}(\Lambda)$. The $\gamma_1, \ldots, \gamma_h$ are closed polygons and the $\lambda_1, \ldots, \lambda_k$ are not closed. The example in Fig. 6.6.1 has one no closed polygon only (due to the special nature of the boundary condition, "half + and half −").

Clearly the correspondence between $(\gamma_1, \gamma_2, \ldots, \gamma_h, \lambda_1, \ldots, \lambda_k)$ and $\underline{\sigma}$ is, *for a fixed boundary condition*, one-to-one except for the case of the periodic or open boundary conditions, when it is one-to-two. Changing boundary conditions implies changing the set of lines (surfaces) which describe the same spin configuration $\underline{\sigma}$.

A very important property of the above geometric description is that, if $|\gamma|$, $|\lambda|$ denote the length (area) of the lines (surfaces) γ and λ, then the energy of a spin configuration is, in zero field, given by

6.6.1
$$H_\Lambda(\underline{\sigma}) = -J \cdot (\text{number of bonds in } \Lambda) + 2J\Big[\sum_i |\gamma_i| + \sum_j |\lambda_j|\Big]\,. \qquad (6.6.1)$$

This remark easily follows from the fact that each bond b contributing $-J$ to the energy has equal spins at its extremes, while the bonds contributing $+J$

have opposite spins at their extremes and, therefore, are cut by a segment of unit length belonging to some γ_i or λ_j.

If $\mathcal{N}_\Lambda$= (number of bonds in Λ), the partition function becomes (in zero field and with fixed spin boundary conditions)

6.6.2
$$Z_\Lambda(\beta) = \sum_{\gamma_1 \ldots \gamma_h} \sum_{\lambda_1 \ldots \lambda_k} \left(e^{-2\beta J \sum_i |\gamma_i|} \cdot e^{-2\beta J \sum_j |\lambda_j|} \right) \cdot e^{\beta J \mathcal{N}_\Lambda} \tag{6.6.2}$$

where the sum runs over the set of lines associated with a spin configuration $\underline{\sigma} \in \mathcal{U}(\Lambda)$ and with the boundary condition under consideration.

In the case of periodic or open boundary conditions there may be no λ's (this happens in the periodic case) and there is an extra factor 2 (because in this case the correspondence between $\underline{\sigma}$ and $(\gamma_1, \ldots, \gamma_h)$ is two-to-one); in the periodic case:

6.6.3
$$Z_\Lambda(\beta) = 2 \sum_{\gamma_1 \cdots \gamma_h} e^{-2\beta J \sum_i |\gamma_i|} \cdot e^{\beta J \mathcal{N}_\Lambda} \tag{6.6.3}$$

and $\mathcal{N}_\Lambda = 2L^2$.

Form the above considerations we draw two important consequences:

(I) If the boundary condition is fixed, the probability of a spin configuration $\underline{\sigma}$ described by $\gamma_1, \ldots, \gamma_h, \lambda_1, \ldots, \lambda_k$ is proportional to:

6.6.4
$$e^{-2\beta J \left(\sum_i |\gamma_i| + \sum_j |\lambda_j| \right)} \tag{6.6.4}$$

(II) In the case of $(+)$ or $(-)$ boundary conditions and two dimensions we remark that $\sum_{\gamma_1 \ldots \gamma_k}$ in (6.6.2) is a sum over "multi-polygons" lying on a shifted lattice and in a box Λ' containing $(L+1)^2$ spins (see the definition in §1.6) and, therefore, if $\sum_i |\gamma_i| = k$ we have

6.6.5
$$Z_\Lambda(\beta) = e^{[2L(L+1)\beta J]} \sum_{k \geq 0} P_k(\Lambda')\, e^{-2\beta J k} \tag{6.6.5}$$

where $P_k(\Lambda')$ is the number of different multi-polygons with perimeter k (see (6.4.12)).

If we now define β^* through

6.6.6
$$\tanh \beta J = e^{-2\beta^* J} \tag{6.6.6}$$

with Λ replaced by a volume Λ' with side $L-1$ then a comparison between (6.6.5) and (6.4.12) yields

6.6.7
$$\frac{Z_\Lambda(\beta)}{2^{L^2} (\cosh \beta J)^{2L(L-1)}} = \frac{Z_{\Lambda'}(\beta^*)}{e^{2\beta^* J L(L-1)}} \tag{6.6.7}$$

Here $Z_\Lambda(\beta)$ is computed with open boundary conditions, while $Z_{\Lambda'}(\beta^*)$ is computed with (+)-boundary conditions.

If we assume that the bulk free energy $f(\beta) = \lim_{\Lambda\to\infty} \frac{1}{|\Lambda|} \log Z_\Lambda(\beta)$ has one and only one singularity as a function of β, for β real, then (6.6.7) can be used to locate the singularity. In fact this implies

6.6.8
$$f(\beta) - \log 2(\cosh \beta J)^2 = f(\beta^*) - 2\beta^* J \tag{6.6.8}$$

having used the fact that the free energy is boundary-condition independent, see (6.5.3). Hence a singularity in β, *if unique*, can take place only when $\beta = \beta^*$, *i.e.* for $\beta = \beta_{c,O}$ such that:

6.6.9
$$\tanh \beta_{c,O} = e^{-2\beta_{c,O} J} \tag{6.6.9}$$

which, indeed, has been shown by Onsager, [On44], to be the exact value of the critical temperature defined as the value of β where $f(\beta)$ is singular (in the sense that its derivative diverges).[12]

In the next section we outline the theory of phase transitions in the Ising model as a macroscopic instability and a spontaneous breakdown of up-down symmetry. We shall concentrate, for geometric reasons, on the two-dimensional Ising model but, unless explicitly stated, the results hold in all dimensions $d \geq 2$.

---

[12] This geometric picture of the spin configurations can be traced back at least as far as Peierls' paper, [Pe36], and has been used, together with formula (6.4.12) to derive (6.6.8) (the "Kramers-Wannier duality" relation) and (6.6.9), [KW41]). A recent interesting generalization of the duality concept has been given in [We71], where some very interesting applications can be found as well as references to earlier works. The duality relation between (+) or (−) boundary conditions and open boundary conditions (which is used here) has been realized by several people. The reader can find other similar interesting relations in [BJS72] and further applications came in [BGJS73]. Duality has found many more applications, see for instance [GHM77] and, for a recent one, [BC94]. In particular a rigorous proof of the correctness of the Onsager-Yang value of the spontaneous magnetization is derived in [BGJS73].

§6.7. Phase Transitions. Existence

In this section we shall show that the (+)-boundary conditions and the (−)-boundary conditions (see §6.3) produce, if the temperature is low enough, different equilibrium states (see §6.3), *i.e.* for large β the correlation functions are different and the difference does not vanish in the limit $\Lambda \to \infty$ (see (6.5.1)).

More precisely we shall prove that if $h = 0$ and β is large enough then

6.7.1 $$\lim_{\Lambda\to\infty} \langle\sigma_x\rangle_{\Lambda,\pm} = \pm m^*(\beta) \neq 0 \tag{6.7.1}$$

where the index $\pm$ refers to the boundary conditions.

Clearly (6.7.1) shows that the magnetization is unstable (in zero field and at low temperature) with respect to boundary perturbations. We also remark that by using periodic boundary conditions one would obtain still another result:

6.7.2 $$\lim_{\Lambda\to\infty} \langle\sigma_x\rangle_{\Lambda,\,periodic} = 0, \qquad \text{if } h = 0 \tag{6.7.2}$$

because $\langle\sigma_x\rangle_{\Lambda,\,periodic} \equiv 0$, if $h = 0$, for obvious symmetry reasons.

After a description of the very simple and instructive proof of (6.7.1) we shall go further and discuss more deeply the character of the phase transition.

As already remarked, spin configurations $\underline{\sigma} \in \mathcal{U}(\Lambda)$ are described in terms of closed polygons $(\gamma_1, \gamma_2, \ldots, \gamma_n)$ if the boundary condition is (+) or (−) and the probability of a configuration $\underline{\sigma}$ described by $(\gamma_1, \gamma_2, \ldots, \gamma_n)$ is proportional to (see (6.6.4)):

6.7.3 $$e^{-2\beta J \sum_i |\gamma_i|}\,. \tag{6.7.3}$$

Below we identify $\underline{\sigma}$ with $(\gamma_1, \gamma_2, \ldots, \gamma_n)$ (with the boundary condition fixed).

Let us estimate $\langle\sigma_x\rangle_{\Lambda,+}$. Clearly $\langle\sigma_x\rangle_{\Lambda,+} = 1 - 2\,P_{\Lambda,+}(-)$, where $P_{\Lambda,+}(-)$ is the probability that in the site x the spin is -1.

We remark that if the site x is occupied by a negative spin then the point x is *inside* some contour γ associated with the spin configuration $\underline{\sigma}$ under consideration. Hence if $\rho(\gamma)$ is the probability that a given contour belongs to the set of contours describing a configuration $\underline{\sigma}$, we deduce

6.7.4 $$P_{\Lambda,+}(-) \le \sum_{\gamma o x} \rho(\gamma) \tag{6.7.4}$$

where $\gamma o x$ means that γ "surrounds" x.

Let us now estimate $\rho(\gamma)$: if $\Gamma = (\gamma_1, \ldots, \gamma_n)$ is a spin configuration and if the symbol $\Gamma\, comp\, \gamma$ means that the contour γ is "disjoint" from (or "compatible" with) $\gamma_1, \ldots, \gamma_n$ (*i.e.* $\{\gamma \cup \Gamma\}$ is a new spin configuration), then

6.7.5 $$\rho(\gamma) = \frac{\sum_{\Gamma\ni\gamma} e^{-2\beta J \sum_{\gamma'\in\Gamma} |\gamma'|}}{\sum_{\Gamma} e^{-2\beta J \sum_{\gamma'\in\Gamma} |\gamma'|}} \equiv e^{-2\beta J|\gamma|} \frac{\sum_{\Gamma\, comp\, \gamma} e^{-2\beta J \sum_{\gamma'\in\Gamma} |\gamma'|}}{\sum_{\Gamma} e^{-2\beta J \sum_{\gamma'\in\Gamma} |\gamma'|}} \tag{6.7.5}$$

Before continuing the analysis let us remark that if $\underline{\sigma} = (\gamma, \gamma_1, \gamma_2, \ldots, \gamma_n)$ then $\underline{\sigma}' = (\gamma_1, \gamma_2, \ldots, \gamma_n)$ is obtained from $\underline{\sigma}$ by reversing the sign of the spins inside γ; this can be used to build an intuitive picture of the second equation in (6.7.5). Clearly the last ratio in (6.7.5) does not exceed 1; hence:

6.7.6
$$\rho(\gamma) \le e^{-2\beta J|\gamma|} \tag{6.7.6}$$

Letting $p = |\gamma|$ and observing that there are at most 3^p different shapes of γ with perimeter p and at most p^2 congruent γ's containing (in their interior) x, we deduce from (6.7.4), (39.6):

6.7.7
$$P_{\Lambda,+}(-) \le \sum_{p=4}^{\infty} p^2 3^p e^{-2\beta J p}\,. \tag{6.7.7}$$

Hence if $\beta \to \infty$ (*i.e.* the temperature $T \to 0$) this probability can be made as small as we like and, therefore, $\langle\sigma_x\rangle_{\Lambda,+}$ is as close to 1 as we like provided β is large enough. It is of fundamental importance that the closeness of $\langle\sigma_x\rangle_{\Lambda,+}$ to 1 is both x and Λ independent.

A similar argument for the $(-)$-boundary condition, or the remark that $\langle\sigma_x\rangle_{\Lambda,-} = -\langle\sigma_x\rangle_{\Lambda,+}$, allows us to conclude that, at large β, $\langle\sigma_x\rangle_{\Lambda,-} \ne \langle\sigma_x\rangle_{\Lambda,+}$ and the difference between the two quantities is uniform in Λ.

Hence we have completed the proof ("*Peierls' argument*") of the fact that there is a strong instability with respect to the boundary conditions of some correlation functions.[13]

We can look upon the above phenomenon as a spontaneous breakdown of up-down symmetry: the Hamiltonian of the model is symmetric, in a zero field, with respect to spin reversal if one neglects the boundary terms; the phase transition manifests itself in the fact that there are equilibrium states in which the symmetry is violated "only on the boundary" and which are not symmetric even in the limit when the boundary recedes to infinity.

§6.8. Microscopic Description of the Pure Phases

The description of the phase transition presented in §6.7 can be made more precise from the physical point of view as well as from the mathematical point of view. A deep and physically clear description of the phenomenon is provided by the theorem below, which also makes precise some ideas familiar from a model, which we shall not discuss here, but which plays an important role in the development of the theory of phase transitions: namely the droplet model, [Fi67c].

Assume that the boundary condition is the $(+)$-boundary condition and describe a spin configuration $\underline{\sigma} \in \mathcal{U}(\Lambda)$ by means of the associated closed disjoint polygons $(\gamma_1, \ldots, \gamma_n)$.

---

[13] The above proof is due to R.B. Griffiths and, independently, to R.L. Dobrushin and it is a mathematically rigorous version of [Pe36].

We regard the ensemble $\mathcal{U}(\Lambda)$ as equipped with the probability distribution attributing to $\underline{\sigma} = (\gamma_1, \ldots, \gamma_n)$ a probability proportional to (6.7.3).
Then the following theorem holds:

Theorem. *If β is large enough there exist $C > 0$ and $\rho(\gamma) > 0$ with $\rho(\gamma) \le e^{-2\beta J|\gamma|}$ and such that a spin configuration $\underline{\sigma}$ randomly chosen out of the ensemble $\mathcal{U}(\Lambda)$ will contain, with probability approaching 1 as $\Lambda \to \infty$, a number $K_{(\gamma)}(\underline{\sigma})$ of contours congruent to γ such that*

6.8.1 $$|K_{(\gamma)}(\underline{\sigma}) - \rho(\gamma)\,|\Lambda|| \le C\sqrt{|\Lambda|}\,e^{-\beta J|\gamma|} \tag{6.8.1}$$

and this relation holds simultaneously for all γ's. In three dimensions one would have $|\Lambda|^{2/3}$ instead of $\sqrt{|\Lambda|}$.

It is clear that the above theorem means that there are very few contours (and that the larger they are the smaller is, in absolute and relative value, their number). The inequality (6.8.1) also implies that for some $C(\beta)$ there are *no* contours with perimeter $|\gamma| > C(\beta)\log|\Lambda|$ (with probability approaching 1 as $\Lambda \to \infty$): this happens when $\rho(\gamma)|\Lambda| < 1$ (because $K_{(\gamma)}(\underline{\sigma})$ is an integer and the right-hand side of (6.8.1) is < 1). Hence a typical spin configuration in the grand canonical ensemble with (+)-boundary conditions is such that the large majority of the spins is "positive" and, in the "sea" of positive spins, there are a few negative spins distributed in small and rare regions (in a number, however, still of order of $|\Lambda|$).
Another nice result which follows from the results of §6.7, and from some improvement, [BS67], of them, concerns the behavior of the equation of state near the phase transition region at low (enough) temperatures.

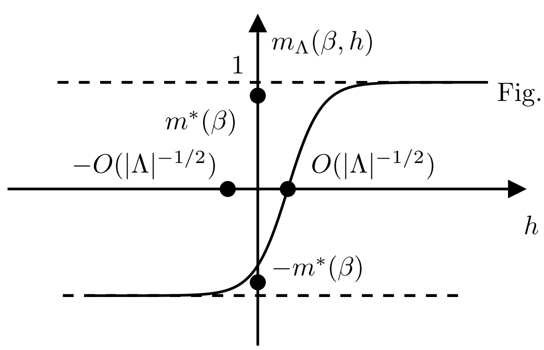


Fig. 6.8.1

If Λ is finite the graph of $h \to m_\Lambda(\beta, h)$f will have a rather different behavior depending on the possible boundary conditions; *e.g.* if the boundary condition is $(-)$ or $(+)$ one gets respectively the results depicted in Fig. 6.8.1 and Fig. 6.8.2, where $m^*(\beta)$ denotes the spontaneous magnetization $\lim_{h\to 0^+} \lim_{\Lambda\to\infty} m_\Lambda(\beta, h)$.

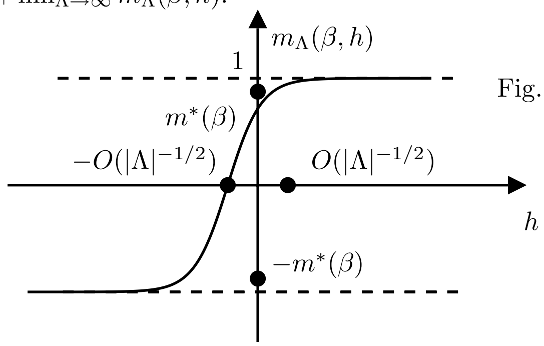


Fig. 6.8.2

With periodic boundary conditions the diagram changes as in Fig. 6.8.3.

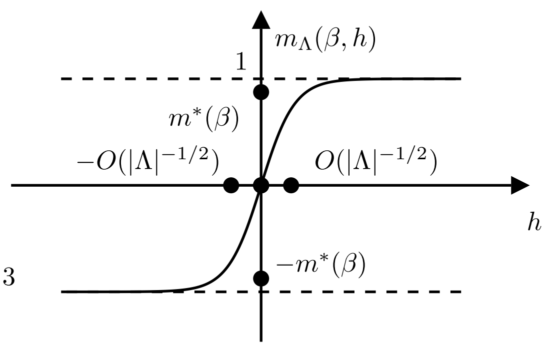


Fig. 6.8.3

The thermodynamic limit $m(\beta, h) = \lim_{\Lambda\to\infty} m_\Lambda(\beta, h)$ exists for all $h \neq 0$ and the resulting graph is as shown in Fig. 6.8.4.

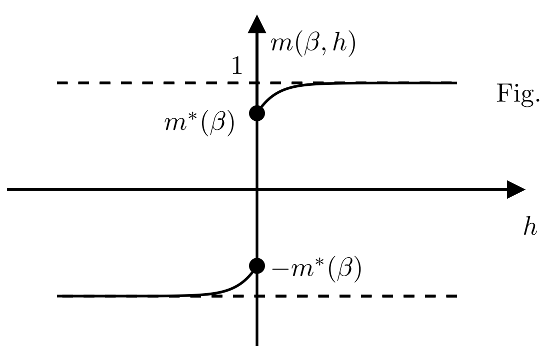


Fig. 6.8.4

At $h = 0$ the limit is not well defined and it depends on the boundary condition (as it must). It can be proven, if β is large enough, that

6.8.2
$$\lim_{h\to 0^+} \frac{\partial\, m(\beta, h)}{\partial h} = \chi(\beta) \tag{6.8.2}$$

is a finite number (*i.e.* the angle between the vertical part of the graph and the rest is sharp, [BS67]).

The above considerations and results also provide a clear idea of what a phase transition for a finite system means.

It is often stated that a finite system "does not" show "sharp" phase transitions; however this statement is always made when considering a fixed boundary condition, usually of periodic or perfect-wall type. By taking into account the importance of the boundary terms we see which kind of phenomena occur in a finite system, if the corresponding infinite system has a sharp phase transition.

The next section is devoted to the discussion of a number of problems concerning the generality of the definition of a phase transition as an instability with respect to the boundary perturbations, and other related problems, in the special case of the Ising model that we are discussing.

§6.9. Results on Phase Transitions in a Wider Range of Temperature

An unpleasant limitation of the results discussed above is the condition of low temperature ("β large enough"). The results of the preceding sections show that, at a low enough temperature, the Ising model is unstable with respect to changes in the boundary conditions. A natural question is whether one can go beyond the low-temperature region and fully describe the phenomena in the region where the instability takes place and first develops. In the particular case of two dimensions it would also be natural to ask whether the maximum value of β to which an instability is associated is the one given by (6.6.9) which corresponds to the value of β where the infinite volume free energy $f(\beta)$ has a singularity, the *critical point*.

The above types of questions are very difficult and are essentially related to the already mentioned theory of the phase transitions based on the search and study of analytic singularities of the thermodynamic functions (which is a theory, however, that has still to be really developed).

Nevertheless a number of interesting partial results are known, which considerably improve the picture of the phenomenon of the phase transitions emerging from the previous sections. A list of such results follows:

(1) It can be shown that the zeros of the polynomial in $z = e^{\beta h}$ given by the product of $z^{|\Lambda|}$ times the partition function (6.2.4) with periodic or perfect-wall boundary conditions lie on the unit circle: $|z| = 1$ ("*Lee-Yang's theorem*"). It is easy to deduce, with the aid of Vitali's convergence theorem for equibounded analytic functions, that this implies that the only singularities of $f(\beta, h)$ in the region $0 < \beta < \infty$, $-\infty < h < +\infty$ can be found at $h = 0$.

A singularity appears if and only if the point $z = 1$ is an accumulation point of the limiting distribution (as $\Lambda \to \infty$) of the zeros on the unit circle.

In fact, if the zeros in question are $z_1, \ldots, z_{2|\Lambda|}$ then

6.9.1
$$\frac{1}{|\Lambda|}\log z^{|\Lambda|} Z(\beta,h,\Lambda,\text{periodic}) = 2\beta J + \beta h + \frac{1}{|\Lambda|}\sum_{i=1}^{2|\Lambda|}\log(z - z_i) \tag{6.9.1}$$

and if $|\Lambda|^{-1}\cdot\big($number of zeros of the form $z_j = e^{i\vartheta_j}$ with $\vartheta \le \vartheta_j \le \vartheta + d\vartheta\big) \xrightarrow[\Lambda\to\infty]{} \rho_\beta(\vartheta)\frac{d\vartheta}{2\pi}$ in a suitable sense, we get from (6.9.1),

6.9.2
$$\beta f(\beta,h) = 2\beta J + \beta h + \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(z - e^{i\vartheta})\,\rho_\beta(\vartheta)d\vartheta \tag{6.9.2}$$

where the second term comes from the $|z|^{|\Lambda|}$ appearing in (6.9.1).

The existence of the measure $\rho_\beta(\vartheta)\frac{d\vartheta}{2\pi}$ such that (6.9.2) is true follows, after some thought, from the existence of the thermodynamic limit $\lim_{\Lambda\to\infty} f_\Lambda(\beta,h) = f(\beta,h)$, [14]

(2) It can be shown that the zeros of the partition function do not move too much under small perturbations of the spin-spin potential even if one allows "many spin" interactions; *i.e.* even if one perturbs the Hamiltonian (6.2.1) with perfect-wall boundary conditions into

6.9.3
$$\begin{aligned} H'_\Lambda(\underline{\sigma}) =& H_\Lambda(\underline{\sigma}) + (\delta H_\Lambda)(\underline{\sigma}) \\ (\delta H_\Lambda)(\underline{\sigma}) =& \sum_{k\ge 1}\sum_{x_1,\ldots,x_k\in\Lambda}\frac{1}{k!}J'(x_1,\ldots,x_k)\,\sigma_{x_1}\ldots\sigma_{x_k} \end{aligned} \tag{6.9.3}$$

where $J'(X)$ is a function of the set $X = (x_1,\ldots,x_k)$ such that

6.9.4
$$||J'|| = \sup_{y\in Z^d}\sum_{y\in X}|J'(X)| \tag{6.9.4}$$

is small enough.

More precisely, suppose that one knows that, when $J' = 0$, the zeros of the partition function in the variable $z = e^{\beta h}$ lie in a certain closed set N of the z-plane. Then if $J' \ne 0$ they lie in a closed set N^1 contained in a neighborhood of N which can be made as small as we please when $||J'|| \to 0$.

This result, [Ru73b], allows us to make a connection between the analyticity properties and the boundary condition instability as described in (3) below.

[14] Here the symbol $\rho_\beta(\vartheta)d\vartheta/2\pi$ has not to be taken too seriously; it really denotes a measure on the circle and this measure is not necessarily $d\vartheta$-continuous. Also the "convergence" statement really means the existence of a measure such that (6.9.2) holds for all real z. The original proof of this theorem is in [LY52]. A much stronger and general statement, the *Ruelle's theorem*, leading in particular to the *Lee-Yang's theorem* is in [Ru71a]: it has been one of the most remarkable among a series of improvements and generalizations of Lee-Yang's theorem (among which I quote [As70], [Ru71b]).

(3) There can be a boundary condition instability only in zero field and, in this case, if and only if the spectrum $\rho_\beta(\vartheta)$ *does not vanish* around $\vartheta = 0$: one says that there is a *gap* around 0 if $\rho_\beta(\vartheta) = 0$ near $\vartheta = 0$.

The proof of this result relies upon 2) and the remark that the correlation functions are functional derivatives with respect to $J'(x_1, \ldots, x_k)$ of the free energy defined by the Hamiltonian (6.9.3), [Ru73b].

(4) Another question is whether the boundary condition instability is always revealed by the one-spin correlation function (as in §6.7) or whether it might be shown only by some correlation functions of higher order. This question is answered by the following result.

There can be a boundary condition instability (at $h = 0$ and β fixed) if and only if

6.9.5
$$\lim_{h\to 0^-} m(\beta, h) \neq \lim_{h\to 0^+} m(\beta, h) \tag{6.9.5}$$

Note that, in view of what was said above (point 3)), $m(\beta, h) = \lim_{\Lambda\to\infty} m_\Lambda(\beta, h)$ is boundary condition independent as long as $h \neq 0$.

In other words there is a boundary condition instability if and only if there is spontaneous magnetization. This rules out the possibility that the phase transition could manifest itself through an instability of some higher-order correlation function which, practically, might be unobservable from an experimental point of view [ML72].

(5) Point (4) implies that a natural definition of the critical temperature T_c is to say that it is the least upper bound of the T's such that (6.9.5) is true ($T = \beta^{-1}$). It is clear that, at this temperature, the gap around $\vartheta = 0$ closes and the function $f(\beta, h)$ has a singularity at $h = 0$ for $\beta > \beta_c = T_c^{-1}$. It can in fact be proven that if (6.9.5) is true for a given β_0 then it is true for all $\beta > \beta_0$, [Gr67], [Fi65].

(6) The location of the singularities of $f(\beta, 0)$ as a function of β remains an open question for $d = 3$, see however [Gr67],[Fi65]. In particular the question of whether there is a singularity of $f(\beta, 0)$ at $\beta = \beta_c$ is open. The identity $\beta_c \equiv \beta_{c,0}$ for the two-dimensional Ising model has been proved in [BGJS73] and, independently, in [AM73].

(7) Finally another interesting question can be raised. For $\beta < \beta_c$ we have instability with respect to the boundary conditions (see (6) above): how strong is this instability? In other words, how many "pure" phases can exist?

Our intuition, in the case of the Ising model, suggests that there should be only two phases: the positively magnetized and the negatively magnetized ones.

To answer the above question in a precise way it is necessary to agree on what a pure phase is, [Ru69], p. 161. We shall call "pure phase" an equilibrium state (see footnote 8, §5.5 and (5.9.18)) if it is translationally

invariant and if the correlation functions have a cluster property of the form

6.9.6
$$\langle \sigma_{x_1} \dots \sigma_{x_n} \sigma_{y_1+a} \dots \sigma_{y_m+a} \rangle \xrightarrow[a\to\infty]{} \langle \sigma_{x_1} \dots \sigma_{x_n} \rangle \langle \sigma_{y_1} \dots \sigma_{y_m} \rangle \tag{6.9.6}$$

where convergence is understood in a very weak sense, *i.e.* the weakest sense which still allows us to deduce that the fluctuations of the extensive quantities are $o(|\Lambda|)$, [Fi65], which is

6.9.7
$$\frac{1}{|\Lambda|} \sum_{a\in\Lambda} \langle \sigma_{x_1} \dots \sigma_{x_n} \sigma_{y_1+a} \dots \sigma_{y_m+a} \rangle \xrightarrow[\Lambda\to\infty]{} \langle \sigma_{x_1} \dots \sigma_{x_n} \rangle \langle \sigma_{y_1} \dots \sigma_{y_m} \rangle \tag{6.9.7}$$

i.e. the convergence in (6.9.6) takes place in the "*Cesaro's limit*" sense.

It can be proved that, in the case of the Ising model, the two states obtained as limits for $\Lambda \to \infty$ of finite volume states (see §6.3) corresponding to (+)-boundary conditions or (−)-boundary conditions are different for $\beta > \beta_c$ and are pure phases in the sense of (6.9.7) above.[15]

Actually it can be proved that, in this case, the limits (6.9.6) exist in the ordinary sense, [GMM72], rather than in the Cesaro sense, and that at low temperature they are approached exponentially fast, see [MS67].

Furthermore, if β is large enough (*e.g.* in two dimension $\sim 10\%$ larger than β_c), these two pure phases exhaust the set of pure phases [GM72a], [Ma72]. For β close to β_c, however, the question is much more difficult: nevertheless it has been completely solved in a remarkable series of papers based on the key work [Ru79b]; see [Hi81], [Ai80]. The work [Ru79b] did provide a real breakthrough and a lot of new ideas for the theory of the Ising model and percolation theory, [Ru81], [Hi97]. The solution of this problem has led to the introduction of many new ideas and techniques in statistical mechanics and probability theory.

Another approach, very rich in results, to the theory of correlation functions originates from the combination of the Griffiths, FKG and other inequalities, see [Gr67], [FKG71], [Le74], with the *infrared bounds* introduced in the work of Mermin and Wagner, [MW66], see §5.8. I only quote here the work [Fr81], mentioned in §5.8 of Chap.V, where the reader can find a very interesting analysis of the behavior at the critical point of various correlations and a clear discussion of the relevance of the dimension of the lattice (if the dimension is ≥ 5 the correlations are "trivial").

Having discussed some exact results about the structure of the phase transition and the nature of pure phases, we shall turn in the next section to the phenomenon of the coexistence of two pure phases.

§6.10. Separation and Coexistence of Pure Phases. Phenomenological Considerations

Our intuition about the phenomena connected with the classical phase transitions is usually based on the properties of the liquid-gas phase transition; this transition is experimentally investigated in situations in which the

[15] This is an unpublished result of R.B. Griffiths. His proof is reported in [GMM72].

total number of particles is fixed (canonical ensemble) and in the presence of an external field (gravity).

The importance of such experimental conditions is obvious; the external field produces a nontranslationally invariant situation and the corresponding separation of the two phases. The fact that the number of particles is fixed determines, on the other hand, the fraction of volume occupied by each of the two phases. The phenomenon of phase transitions in the absence of an external field will be briefly discussed in §6.14.

In the framework of the Ising model it will be convenient to discuss the phenomenon of phase coexistence in the analog of the canonical ensemble $\mathcal{U}(\Lambda, m)$, introduced and discussed in §6.2, where the total magnetization $M = m|\Lambda|$ is held fixed.

To put ourselves in the phase transition region we shall take β large enough and, for a fixed α, $0 < \alpha < 1$:

6.10.1
$$m = \alpha\, m^*(\beta) + (1-\alpha)\,(-m^*(\beta)) = \,(1-2\alpha)\, m^*(\beta) \qquad (6.10.1)$$

i.e. we put ourselves in the vertical "plateau" of the diagram $(m, h)_\beta$ (see Fig. 6.8.4 above).

Fixing m as in (6.10.1) does not yet determine the separation of the phases in two different regions; to obtain this effect it will be necessary to introduce some external cause favoring the occupation of a part of the volume by a single phase. Such an asymmetry can be obtained in at least two ways: through a weak uniform external field (in complete analogy with the gravitational field in the liquid-vapor transition) or through an asymmetric field acting only on the boundary spins. This second way should have the same qualitative effect as the former, because in a phase transition region a boundary perturbation produces volume effects (this last phenomenon, which has been investigated in the previous sections, is often also referred to as the "*long-range order*" of the correlations).

From a mathematical point of view it is simple to use a boundary asymmetry to produce phase separations.

To obtain a further, but not really essential, simplification of the problem consider the two-dimensional Ising model with $(+,-)$-cylindrical or $(+,+)$-cylindrical boundary conditions.

The spins adjacent to the bases of Λ act as symmetry-breaking external fields. The $(+,+)$-cylindrical boundary condition should, clearly, favor the formation inside Λ of the positively magnetized phase; therefore it will be natural to consider, in the canonical ensemble, this boundary condition only when the total magnetization is fixed to be $+m^*(\beta)$ (see Fig. 6.8.4).

On the other hand, the boundary condition $(+,-)$ favors the separation of phases (positively magnetized phase near the top of Λ and negatively magnetized phase near the bottom).

Therefore it will be natural to consider this boundary condition in the case of a canonical ensemble with magnetization $m = (1-2\alpha)\, m^*(\beta)$ with $0 < \alpha < 1$, (6.10.1).

In this last case one expects, as already mentioned, the positive phase to adhere to the top of Λ and to extend, in some sense to be discussed, up to a distance $O(L)$ from it; and then to change into the negatively magnetized pure phase.

To make precise the above phenomenological description we shall describe the spin configurations $\underline{\sigma} \in \mathcal{U}(\Lambda, m)$ through the associated sets of disjoint polygons (cf. §6.6).

Fix the boundary conditions to be $(+,+)$ or $(+,-)$-cylindrical boundary conditions and note that the polygons associated with a spin configuration $\underline{\sigma} \in \mathcal{U}(\Lambda, m)$ are *all closed* and of two types: the ones of the first type, denoted $\gamma_1, \ldots, \gamma_n$, are polygons which do not encircle Λ, the second type of polygons, denoted by the symbols λ_α, are the ones which wind up, at least once, around Λ.

So a spin configuration $\underline{\sigma}$ will be described by a set of polygons $(\gamma_1, \ldots, \gamma_n, \lambda_1, \ldots, \lambda_h)$. It is, perhaps, useful to remark once more that the configuration $\underline{\sigma}$ will be described by different sets of polygons according to which boundary condition is used (among the ones we are considering, *i.e.* $(+,+)$ or $(+,-)$-boundary conditions). However, for a fixed boundary condition, the correspondence between spin configuration and sets of disjoint closed contours is one-to-one and the statistical weight of a configuration $\underline{\sigma} = (\gamma_1, \ldots, \gamma_n, \lambda_1, \ldots, \lambda_h)$ is (cf. (6.6.4)):

6.10.2
$$e^{-2\beta J(\sum_i |\gamma_i| + \sum_j |\lambda_j|)} \,. \tag{6.10.2}$$

It should also be remarked that the above notation is not coherent with the notation of §6.6, where the symbol λ is used for open polygons (absent here); but this will not cause any confusion. The reason why we call λ the contours that go around the cylinder Λ is that they "look like" open contours if one forgets that the opposite sides of Λ have to be identified.

It is very important to remark that if we consider the $(+,-)$-boundary conditions then the number of polygons of λ-type must be *odd* (hence $\neq 0$), while if we consider the $(+,+)$-boundary condition then the number of λ-type polygons must be *even* (hence it could be 0).

§6.11. Separation and Coexistence of Phases. Results

Bearing in mind the geometric description of the spin configuration in the canonical ensembles considered with the $(+,+)$-cylindrical or the $(+,-)$-cylindrical boundary conditions (which we shall denote briefly as $\mathcal{U}^{++}(\Lambda, m), \mathcal{U}^{+-}(\Lambda, m)$) we can formulate the following theorem, [GM72b], essentially developed by Minlos and Sinai to whom the very foundations of the microscopic theory of coexistence is due:

Theorem. *For $0 < \alpha < 1$ fixed let $m = (1 - 2\alpha)\, m^*(\beta)$; then for β large enough a spin configuration $\underline{\sigma} = (\gamma_1, \ldots, \gamma_n, \lambda_1, \ldots, \lambda_{2h+1})$ randomly chosen out of $\mathcal{U}^{+-}(\Lambda, m)$ enjoys the properties (1)-(4) below with a probability (in $\mathcal{U}^{+-}(\Lambda, m)$) approaching 1 as $\Lambda \to \infty$:*

(1) $\underline{\sigma}$ contains only one contour of λ-type and

6.11.1 $$|\,|\lambda| - (1+\varepsilon(\beta))L| < o(L) \qquad (6.11.1)$$

where $\varepsilon(\beta) > 0$ is a suitable (α-independent) function of β tending to zero exponentially fast as $\beta \to \infty$.

(2) If $\Lambda_\lambda, \Lambda'_\lambda$ denote the regions above and below λ we have

6.11.2 $$|\,|\Lambda_\lambda| - \alpha\,|\Lambda|\,| < \kappa(\beta)\,|\Lambda|^{3/4} \qquad (6.11.2)$$

6.11.3 $$|\,|\Lambda'_\lambda| - (1-\alpha)|\Lambda|\,| < \kappa(\beta)\,|\Lambda|^{3/4} \qquad (6.11.3)$$

where $\kappa(\beta) \xrightarrow[\beta\to\infty]{} 0$ exponentially fast.

(3) If $M_\lambda = \sum_{x\in\Lambda_\lambda} \sigma_x$, we have

6.11.4 $$|M_\lambda - \alpha\, m^*(\beta)\,|\Lambda|| < \kappa(\beta)|\Lambda|^{3/4} \qquad (6.11.4)$$

and a similar inequality holds for $M'_\lambda = \sum_{x\in\Lambda'_\lambda} \sigma_x = m|\Lambda| - M_\lambda$.

(4) If $K^\lambda_\gamma(\underline{\sigma})$ denotes the number of contours congruent to a given γ and lying in Λ_λ then, simultaneously for all the shapes of γ:

6.11.5 $$|\,K^\lambda_\gamma(\underline{\sigma}) - \rho(\gamma)\,\alpha\,|\Lambda|\,| \le Ce^{-\beta J|\gamma|}\sqrt{|\Lambda|} \qquad C > 0 \qquad (6.11.5)$$

*where $\rho(\gamma) \le e^{-2\beta J|\gamma|}$ is the same quantity already mentioned in the text of the theorem of §6.8. A similar result holds for the contours below λ (*cf.* the comments on (6.8.1)).*

It is clear that the above theorem not only provides a detailed and rather satisfactory description of the phenomenon of phase separation, but also furnishes a precise microscopic definition of the line of separation between the two phases, which should be naturally identified with the (random) line λ.

A very similar result holds in the ensemble $\mathcal{U}^{++}(\Lambda, m^*(\beta))$: in this case 1) is replaced by

(1') *no* λ-type polygon is present

while (2), (3) become superfluous and 4) is modified in the obvious way. In other words a typical configuration in the ensemble $\mathcal{U}^{++}(\Lambda, m^*(\beta))$ has the same appearance as a typical configuration of the grand canonical ensemble $\mathcal{U}(\Lambda)$ with (+)-boundary condition (which is described by the theorem of §6.8).

We conclude this section with a remark about the condition that $0 < \alpha < 1$ has to be fixed beforehand in formulating the above theorem. Actually the results of the theorem hold at fixed β (small enough) for all the α's such

that $\varepsilon(\beta) < \min(\alpha, 1-\alpha)$, *i.e.* such that the line λ cannot touch the bases of Λ (in which case there would be additional physical phenomena and correspondingly different results).

§6.12. Surface Tension in Two Dimensions. Alternative Description of the Separation Phenomena

A remarkable application of the above theorem is the possibility of giving a microscopic definition of surface tension between the two pure phases, [GM72b]. We have seen that the partition functions

6.12.1
$$Z^{++}(\Lambda,\beta) = \sum_{\underline{\sigma}\in\mathcal{U}^{++}(\Lambda,m^*(\beta))} e^{-2\beta J(\sum_i |\gamma_i| + \sum_j |\lambda_j|)} \tag{6.12.1}$$

and (if $m = (1-2\alpha)m^*(\beta)$, $0 < \alpha < 1$)

6.12.2
$$Z^{+-}(\Lambda,\beta) = \sum_{\underline{\sigma}\in\mathcal{U}^{+-}(\Lambda,m)} e^{-2\beta J(\sum_i |\gamma_i| + \sum_j |\lambda_j|)} \tag{6.12.2}$$

will essentially differ, at low temperature, only because of the line λ (present in $\mathcal{U}^{+-}(\Lambda,m)$ and absent in $\mathcal{U}^{++}(\Lambda,m^*(\beta))$, see the preceding section).

A natural definition (in two dimensions) of surface tension between the phases, based on obvious physical considerations, can therefore be given in terms of the different asymptotic behavior of $Z^{++}(\Lambda,m^*(\beta))$ (or of the grand canonical $Z^{++}(\Lambda,\beta)$) and $Z^{+-}(\Lambda,m)$:

6.12.3
$$\beta\,\tau(\beta) = \lim_{\Lambda\to\infty} \frac{1}{L}\log\frac{Z^{+-}(\Lambda,m)}{Z^{++}(\Lambda,m^*(\beta))}\,. \tag{6.12.3}$$

The above limit (which should be α-independent for $\varepsilon(\beta) < \min(\alpha,1-\alpha)$, cf. the concluding remarks of the preceding section) can be exactly computed at low enough temperature and is given by

6.12.4
$$\beta\,\tau(\beta) = -2\beta J - \log\tanh\beta J \tag{6.12.4}$$

which is the value computed by Onsager, [On44], by using a different definition, not based on the above detailed microscopic description of the phases and of the line of separation: for a comparison of various old definitions of surface tension, new ones and a proof of their equivalence see [AGM71].

We conclude this section with a brief discussion of one particular but very convenient alternative way of investigating the phenomenon of coexistence of two phases. Another still different way of investigating the phenomenon will be discussed in §6.14.

Consider the grand canonical ensemble, but impose the following boundary conditions: the spins adjacent to the upper half of the boundary of Λ are fixed to be $+1$, while the ones adjacent to the lower half are -1 (and no periodicity condition). This is an $\underline{\varepsilon}$-type boundary condition (see §6.3 and

Fig. 6.6.1, or also cover figure) generating an ensemble that we shall denote by $\mathcal{U}_0^{+-}(\Lambda)$.

It is clear that a configuration $\underline{\sigma} \in \mathcal{U}_0^{+-}(\Lambda)$ is described, under the above boundary condition, by one *single* open polygon λ (surface in three dimensions) going from one side of Λ to the opposite side, and by a set of disjoint closed polygons (polyhedra in three dimensions) $(\gamma_1, \ldots, \gamma_n)$.

The surface λ now plays the role of the polygons encircling Λ in the case of cylindrical boundary conditions (and two dimensions) and it is also clear that a theorem very similar to those already discussed should hold in this case. The above point of view is more relevant in the three-dimensional case where a "cylindrical" boundary condition would have a less clear physical meaning, and it would look rather a mathematical device.

In the three-dimensional case λ is a "surface" with a boundary formed by the square in the "middle" of $\partial\Lambda$ where the "break" between the spins fixed to be $+1$ and the ones fixed to be -1 is located.

In the next section we investigate in more detail the structure of such a line or surface of separation between the phases.

§**6.13. The Structure of the Line of Separation. What a Straight Line Really is**

The theorem of §6.11 tells us that, if β is large enough, then the line λ is almost straight (since $\varepsilon(\beta)$ is small). It is a natural question to ask whether the line λ is straight in the following sense: suppose that λ, regarded as a polygon belonging to a configuration $\underline{\sigma} \in \mathcal{U}^{+-}(\Lambda, m)$ (cf. §6.11), passes through a point $q \in \Lambda$; then we shall say that λ is "straight" or "rigid" if the (conditional) probability $\mathcal{P}_\lambda$ that λ passes also through the site q', opposite[3] to q on the cylinder Λ, *does not* tend to zero as $\Lambda \to \infty$, otherwise we shall say that λ is *not rigid* or *fluctuates*. Of course the above probabilities must be computed in the ensemble $\mathcal{U}^{+-}(\Lambda, m)$.

Alternatively (and essentially equivalently) we can consider the ensemble $\mathcal{U}_0^{+-}(\Lambda)$ (see §6.12, *i.e.* the grand canonical ensemble with the boundary condition with the boundary spins set to $+1$ in the upper half of $\partial\Lambda$, vertical sites included, and to -1 in the lower half). We say that λ is rigid if the probability that λ passes through the center of the box Λ (*i.e.* 0) *does not* tend to 0 as $\Lambda \to \infty$; otherwise it is not rigid.

It is rather clear what the above notion of rigidity means: the "excess" length $\varepsilon(\beta)L$, see (6.11.1), can be obtained in two ways: either the line λ is essentially straight (in the geometric sense) with a few "bumps" distributed with a density of order $\varepsilon(\beta)$ or, otherwise, the line λ is bent and, therefore, only locally straight and part of the excess length is gained through the bending.

In three dimensions a similar phenomenon is possible. As remarked at the end of the last section, in the ensemble $\mathcal{U}_0^{+-}(\Lambda)$, in this case λ becomes a

[3] *i.e.* on the same horizontal line and $L/2$ sites apart.

surface with a square boundary fixed at a certain height (*i.e.* 0), and we ask whether the center of the square belongs to λ with non vanishing probability in the limit $\Lambda \to \infty$.

The rigidity or otherwise of λ can, in principle, be investigated by optical means; one can have interference of coherent light scattered by surface elements of λ separated by a macroscopic distance only if λ is rigid in the above sense.

It has been rigorously proved that, at least at low temperature, the line of separation λ is *not rigid* in two dimensions (and the fluctuation of the middle point is of the order $O(\sqrt{|\Lambda|})$); a very detailed description of the separation profile is available, at low temperature ([Ga72a],[GV72],[Ga72b]) and even all the way to the critical point [AR76]. In three dimensions the situation is very different: it has been shown that the surface λ is rigid at low enough temperature, see [Do72], [VB75]. The latter reference provides a very nice and simple argument for the three-dimensional rigidity.

An interesting question remains open in the three-dimensional case and is the following: it is conceivable that the surface, although rigid at low temperature, might become loose at a temperature $\tilde{T}_c$ smaller than the critical temperature T_c (the latter being defined as the highest temperature below which there are at least two pure phases). The temperature $\tilde{T}_c$, if it exists, is called the "*roughening transition*" temperature, see [KM86], [VB77], [KM87], [VN87], .

It would be interesting to examine the available experimental data on the structure of the surface of separation to set limits on $T_c - \tilde{T}_c$ in the case of the liquid-gas phase transition where an analogous phenomenon can conceivably occur even though a theory of it is far from being in sight, at least if one requires a degree of rigor comparable to that in the treatment of the results so far given for the Ising model.

We conclude by remarking that the rigidity of λ is connected with the existence of translationally noninvariant equilibrium states (see §6.3).

The discussed nonrigidity of λ in two dimensions provides the intuitive reason for the absence of nontranslationally invariant states.

Note that the existence of translationally noninvariant equilibrium states is not necessary for the description of coexistence phenomena. The theory of the two-dimensional Ising model developed in the preceding sections is a clear proof of this statement.

§6.14. Phase Separation Phenomena and Boundary Conditions. Further Results

The phenomenon of phase separation described in §6.12 and §6.13 is the ferromagnetic analogue of the phase separation between a liquid and a vapor in the presence of the gravitational field.

It is relevant to ask to what extent an external field (or some equivalent boundary condition) is really necessary; for instance one could imagine a situation in which two phases coexist in the absence of any external field.

Let us discuss first some phenomenological aspects of the liquid-gas phase separation in the absence of external fields. One imagines that, if the density is fixed and corresponds to some value on the "plateau" of the phase diagram, see Fig. 5.1.1, then the space will be filled by vapor and drops of liquid in equilibrium. Note that the drops will move and, from time to time, collide; since the surface tension is negative the drops will tend to cluster together and, eventually, in an equilibrium situation there will be just one big drop (and the drop surface will be minimal). The location of the drop in the box Λ will depend on how the walls are made and how they interact with the particles within Λ.

Let us consider some extreme cases:

(1) the walls "repel" the drops,

(2) the walls "attract" the drops,

(3) the wall is perfect and does not distinguish between the vapor and the liquid.

In the first case the drop will stay away from the boundary $\partial\Lambda$ of Λ. In the second case the drop will spread on the walls, which will be wet as much as possible. In the third case it will not matter where the drop is; the drop will be located in a position that minimizes the "free" part of its boundary (*i.e.* the part of the boundary of the drop not on $\partial\Lambda$). This means that the drop will prefer to stay near a corner rather than wetting all the wall.

Let us translate the above picture into the Ising model case. Assume that β is large and $m = (1-2\alpha)m^*(\beta)$ (see Fig. 6.8.4) (*i.e.* assume that the magnetization is on the vertical plateau of the $(m,h)_\beta$ diagram in Fig. 6.8.4).

Then the conditions (1), (2), (3) can be realized as follows:

(1) The spins adjacent to the boundary are all fixed to be $+1$. This favors the adherence to the boundary of the positively magnetized phase.

(2) The spins adjacent to the boundary are all fixed to be -1. This favors the adherence to the boundary of the negatively magnetized phase.

(3) There are no spins adjacent to the boundary, *i.e.* we consider *perfect wall* or *open* (or *free*) boundary conditions (see §6.3).

The rigorous results available in the case of the Ising model confirm the above phenomenological analysis of liquid-vapor coexistence [MS67]:

Theorem. *Fix $0<\alpha<1$ and consider $(+)$-boundary conditions. Then a spin configuration $\underline{\sigma}$ randomly extracted from the canonical ensemble with magnetization $m = (1-2\alpha)m^*(\beta)$ has, if β is large enough, properties (1)÷(3) below with a probability tending to 1 as $\Lambda\to\infty$.*

(1) There is only one γ such that $|\gamma| > \frac{1}{333}\log|\Lambda|$ and it has the property[16]

6.14.1
$$\big|\,|\gamma| - 4\sqrt{(1-\alpha)|\Lambda|}\,\big| \le \delta(\beta)\sqrt{|\Lambda|} \tag{6.14.1}$$

with $\delta(\beta)\to 0$ as $\beta\to\infty$ (exponentially fast); % (2) The area enclosed by γ is $\vartheta(\gamma)$:

6.14.2
$$\big|\,\vartheta(\gamma) - (1-\alpha)|\Lambda|\,\big| \le \kappa(\beta)\,|\Lambda|^{3/4} \tag{6.14.2}$$

(3) The magnetization $M(\vartheta(\gamma))$ inside γ is on the average equal to $-m^*(\beta)$ and, more precisely,

6.14.3
$$\big|\,M(\vartheta(\gamma)) + m^*(\beta)\,(1-\alpha)\,|\Lambda|\,\big| \le \kappa(\beta)\,|\Lambda|^{3/4} \tag{6.14.3}$$

and, therefore, the average magnetization outside $\vartheta(\gamma)$ is $+m^*(\beta)$.

This theorem also holds in three dimensions but, of course, the exponent of $|\Lambda|$ in (6.14.1) changes (from $\frac{1}{2}$ to $\frac{2}{3}$).

The above theorem shows that a typical configuration consists of a positively magnetized pure phase adherent to the boundary and of a "drop" of negatively magnetized phase *not adhering* to the boundary (since γ is closed). The size of the drop is $\sim\sqrt{(1-\alpha)|\Lambda|}$ (as it should be).

Note that the drop is almost square in shape (as follows from (6.14.1), (6.14.2)): this should not be astonishing since the space is discrete and the isoperimetric problem on a square lattice has the square as a solution (rather than a circle).

The opposite situation is found if one fixes a $(-)$-boundary condition; a square drop forms in the middle of the box with side $\sim\sqrt{\alpha|\Lambda|}$ and average magnetization $m^*(\beta)$.

Finally if the boundary condition is of perfect wall type ($\mathcal{B}_\Lambda(\underline{\sigma})\equiv 0$), then the above theorem does not hold and one can prove (say, in two dimensions) that a typical spin configuration has just one open contour λ (with ends on $\partial\Lambda$) which separates the space in two parts which are occupied by opposite phases; the line λ should be the shortest possible compatible with the condition that the volume Λ is divided by it into two regions of volume essentially $\alpha|\Lambda|$ and $(1-\alpha)|\Lambda|$ (respectively occupied by the positively magnetized phase and by the negatively magnetized phase): see [Ku83].

If one interprets the spins equal to $+1$ as particles and the spins equal to -1 as empty sites, then one has a lattice gas model which undergoes a liquid-vapor phase transition presenting the phenomenological aspects outlined at the beginning of this section for these transitions.

---

[16] The number 333 is just an arbitrary constant and it is reported here because it appeared in the original literature, [MS67], as a joke referring to the contemporary papers on the KAM theorem ("Moser's constant"). In fact it looks today somewhat confusing and quite strange: the modern generation do not seem to appreciate this kind of hum-our any more; they became more demanding and would rather ask here for the "best" constant; this is my case as well.

To conclude we remark that, in the phase separation phenomenon, the finiteness of the box only plays the role of fixing the density. The detailed structure of the phenomenon depends on the boundary conditions which, in experimental situations, turn out to be something intermediate between the three extreme cases discussed above.

Note that (6.14.1) does not provide a satisfactory estimate of $|\lambda|$ since the allowed error is still of the order of $\sqrt{|\Lambda|}$; but better estimates can be obtained to determine *exactly* (*i.e.* with an error much smaller than $O(\sqrt{|\Lambda|})$) the size of the boundary and its macroscopic shape. Such remarkable results also provide a rigorous microscopic theory of the ancient *Wulff's construction* of the shape of a droplet in the absence of gravity when the surface tension is not spherically symmetric as a function of the normal to the droplet surface, see [DKS92], [PV96], [MR94], [Mi95]. The results hold in the two dimensional case and for low enough temperature: their literal extension to the whole coexistence region or to the three-dimensional case (even at very low temperature) seem out of reach, if at all possible, of present day techniques: in this respect, unexpected, remarkable progress has been achieved very recently, [PV99], with new techniques that cover, at least in 2 dimensions the whole phase–coexistence region (showing that despair is out of place). However one can get surprisingly detailed informations by general considerations based on inequalities and convexity properties of the surface tension, see [MMR92].

Another problem is the investigation of the dependence of the correlation functions on the distance from the surface of the drop.

The analogs of the first two questions just raised were previously satisfactorily answered in the two-dimensional Ising model with the "easier" cylindrical boundary conditions (see §6.11), *i.e.* in the case of an "infinite" drop with a flat surface. This problem has been approximately studied even in the case of a flat drop, [BF67].

§6.15. Further Results, Some Comments and Some Open Problems

In §6.14 we dealt with the case of a nearest neighbor Ising model. It has become customary, in the literature, to apply the name of Ising model to more general models in which the "bulk" Hamiltonian (*i.e.* without the boundary interactions and conditions) has the form, see §5.10,

6.15.1
$$-h\sum_{x_i}\sigma_{x_i}-\sum_{i<j}J_2(x_i,x_j)\sigma_{x_i}\sigma_{x_j}-\sum_{i<j<k}J_3(x_i,x_j,x_k)\sigma_{x_i}\sigma_{x_j}\sigma_{x_k}+\ldots \tag{6.15.1}$$

where the potentials $J_n(x_1,\ldots,x_n)$ are translationally invariant functions of $(x_1,\ldots,x_n)$ and satisfy certain restrictions of the type

6.15.2
$$\sum_x |J_2(0,x)|+\sum_{x,y}|J_3(0,x,y)|+\ldots<+\infty\,. \tag{6.15.2}$$

If only pair potentials are present, *i.e.* if the bulk Hamiltonian has the form:

6.15.3
$$-h\sum_x \sigma_x - \sum_{i<j} J(x_i - x_j)\sigma_{x_i}\sigma_{x_j} \tag{6.15.3}$$

and if $J(r) \le 0$, then most of the results described in this chapter and appropriately reformulated can be extended or become very reasonable conjectures, see [Ru69], p. 125, [MS67], [BS67], [Do68c] and the review [Gi70].

Many results remain true for more general pair potentials and for other models (like continuous gases) at least from the qualitative point of view; in fact it is reasonable that the results selected here for discussion should have, at least qualitatively, an analog in the "general" case of a classical (as opposed to quantum) phase transition.

Results such as analyticity and absence of phase transitions at high temperature, or exact solutions, are a peculiarity of lattice models and have been partly discussed in Chap.V and they will be analyzed again later.

Below I list a number of rather randomly chosen and interesting problems suggested by the topics of this chapter.

(1) The solution of the two-dimensional Ising model is based on the so called "transfer matrix". The investigation of the transfer matrix has been pursued in some detail in the case of periodic or open boundary conditions in two or three dimensions, [MS70], [CF71], see also [On44],[Ab71].
The transfer matrix with non-symmetric boundary conditions has also been studied in the two-dimensional case,[Ab84], *i.e.* the transfer matrix between two rows (or planes) where the line (or surface) of separation should pass if straight. A qualitative difference should arise between two and three dimensions (see, for more details, §6.15).

(2) In Fig. 6.8.4 we see that the isotherm $m(\beta,h)$ as a function of $h>0$ abruptly ends at $h=0$. It is a natural question whether $h=0$ is an analytic singularity of $m(\beta,h)$ or whether $m(\beta,h)$ can be analytically continued to $h<0$. There has always been strong evidence for a singularity, [LR69], and it has been shown, rigorously, that at $h=0$ there is an essential singularity, at least at large β, although the function $m(\beta,h)$ is infinitely differentiable as a function of h for $h\ge 0$, [Is84].

(3) The answer to (2) makes clear that one has to give up the theory of "*metastability*" based on the possibility of an analytic continuation of the magnetization as a function of h through $h=0$. The latter idea was founded in fact on the absence of an analytic singularity at $h=0$ in the equation of state deduced from mean field theory (that is in the van der Waals theory of phase transitions, whose version for spin systems is called the *Curie Weiss' theory*): for an interesting mathematically complete treatment of the metastability phenomenon in the case of very weak and very long ranged forces see [LP79]. The question of how one can explain metastability phenomena in systems with short range forces has been investigated in great

detail as a *dynamical phenomenon* and the results are very many, detailed and varied, see for instance [CCO74], [KO93],[MOS90].

(4) There is a great number of other lattice models for which phase transitions are proven to take place; the basic techniques originated by the work of Minlos and Sinai, [MS67], have evolved and have been highly enriched; one can compare the situation in the early 1970s, before the first proof of the existence of a phase transition in a continuous system, still based on symmetry breaking [Ru73b], see the review [Gi70], with the present day very refined results (see [BLPO79], [BC94]. In this respect one must mention the very recent progress in *nonlattice models*, [Jo95] and [MLP98]). The latter works should be considered major breakthroughs on the problem of phase transitions occurring in continuous systems or, more generally, in systems without special symmetries which are spontaneously broken at the transition.

(5) The question of whether, once a phase transition is known to occur, one can count how many pure phases exist is often a very intricate question as various examples show (see [BC94]).

(6) A detailed description of the correlation functions near the line or surface of separation has still to be discussed, see [AR76].

(7) The microscopic definition of surface tension in the particular case of the three-dimensional Ising model has been studied but there are many open problems, see [Mi95], particularly concerning the cases when the boundary conditions would impose an ideal surface of separation, between the two phases, which is not parallel to the lattice planes. Furthermore it is a well founded conjecture that there is a temperature *lower than the critical temperature* for the appearance of spontaneous magnetization, above which the separation surface shows large fluctuations (possibly of order $\sqrt{\log L}$), see [VB77]. In this regime there would probably be no more translationally noninvariant states, and it is likely that the surface tension $\tau(\beta)$ is not analytic as a function of β (while at low temperature it is known that the surface tension relative to an ideal surface of separation parallel to the lattice planes is such that $\tau(\beta) + 2\beta J$ is analytic in $e^{-\beta J}$). This would identify a second type of phase transition which has been called in §6.13 the *roughening transition*, see [KM86], [KM87] and, for a review, [VN87].

(8) The problem of the existence of phase transitions in models close to symmetric models but asymmetric was expected to give interesting results, [Me71]. Substantial progress towards the understanding of phase transitions not directly associated with spontaneous breakdown of symmetry has been achieved by the understanding of the model in (6.15.1) with $J_3 \neq 0$, [PS76]. Although the models in [PS76] are "close" to symmetric models the absence of a rigorous symmetry was a *major obstacle* and the solution proposed in [PS76] has generated a large number of investigations: the theory is

generally known as the *Pirogov-Sinai theory* of phase coexistence, see for instance the applications to Potts' models, [LMMRS91], [MMRS91].

(9) In connection with (8) an interesting problem arises on the correct definition of *approximate symmetry*. The analysis has been attempted in [EG75] but the results are still very partial. There is also the possibility that in some cases symmetries that are apparently already broken in the Hamiltonian are in fact dynamically restored (and then, possibly, spontaneously broken). An example in which this happens has been proposed in [BG97] (see [BGG97] and [Ga98b] for some applications) in a *totally different context*, but its relevance for the theory of phase transitions might create surprises.

(10) Last but not least, the phase transitions problem in quantum statistical mechanics will not be discussed in this book, but it is, of course, very important. The conceptual frame in which it is developed in the literature is the same as the one we described in the classical cadre. However the phenomenology becomes even richer and full of surprises: see [BCS57],[Br65]. Phase transitions of quantum systems can be studied in "lattice systems" at a rather sophisticated mathematical level, [Gi69b], [DLS78], and we shall meet some in Chap.VII where they appear because of their relation with transitions in classical spin models. However the theory is, generally speaking, less developed on a mathematical level but it is enormously developed on a phenomenological level, see [An84].

.

Chapter *VII*

Exactly Soluble Models

.

§7.1. Transfer Matrix in the Ising Model: Results in $d = 1, 2$

Many well-known exact solutions of statistical mechanical models are based on the *transfer matrix* method. In fact the summation over the states of the system, *e.g.* the sum over the values of the spin at each site, can be quite easily interpreted as an operation of summing over the labels of a product of matrices in order to compute the trace of the product. Thus the problem of computing, say, a partition function is "reduced" to that of diagonalizing certain matrices with the purpose of computing their eigenvalues and eigenvectors. The latter sometimes also provide informations on the correlation functions.

The difficulty is that the matrices that one obtains are "large dimension matrices" and they are difficult to diagonalize. This can be done, at the cost of remarkable effort, in a few cases. We shall discuss a few samples of them. The simplest case is the $d = 1$ Ising model already studied in §6.4 with a different method and at zero field h.

Consider the one dimensional Ising model with periodic boundary conditions, see (6.2.1), Ch.IV. If $\sigma_{L+1} \equiv \sigma_1$ the partition function $Z(\Lambda, \beta, h)$ can be written as

7.1.1
$$\begin{aligned}\sum_{\sigma_1\ldots\sigma_L}\prod_{i=1}^{L} e^{\beta J\sigma_i\sigma_{i+1}+\beta h\sigma_i} &= \sum_{\sigma_1\ldots\sigma_L}\prod_{i=1}^{L} e^{\frac{\beta}{2}h\sigma_i} e^{\beta J\sigma_i\sigma_{i+1}} e^{\frac{\beta}{2}h\sigma_{i+1}} = \\ = \sum_{\sigma_1\ldots\sigma_L} V_{\sigma_1\sigma_2}V_{\sigma_2\sigma_3}\ldots V_{\sigma_L\sigma_1} &= \,\mathrm{Tr}\, V^L\end{aligned} \tag{7.1.1}$$

where V is a two-by-two matrix such that $(\sigma, \sigma' = \pm 1)$

7.1.2
$$V_{\sigma\sigma'} = e^{\frac{\beta}{2}h\sigma} e^{\beta J\sigma\sigma'} e^{\frac{\beta}{2}h\sigma'}, \qquad V = \begin{pmatrix} e^{\beta(h+J)} & e^{-\beta J} \\ e^{-\beta J} & e^{-\beta(h-J)}\,. \end{pmatrix} \tag{7.1.2}$$

If $\lambda_+ > \lambda_-$ are the two eigenvalues of V, we find

7.1.3
$$Z(\Lambda, \beta, h) = \lambda_+^L + \lambda_-^L \tag{7.1.3}$$

so that

7.1.4
$$\beta f(\beta, h) = \lim_{L\to\infty} \frac{1}{L}\log Z = \log \lambda_+\,. \tag{7.1.4}$$

It is easy to check that $\lambda_+(\beta, h)$ is analytic in β and h for $0 < \beta < \infty$ and $-\infty < h < \infty$, *i.e.* there are no phase transitions (as singularities of $f(\beta, h)$). In fact

7.1.5
$$\beta f(\beta, h) = \log\Big(e^{\beta J}\cosh\beta h \,+ \big(e^{2\beta J}(\sinh\beta h)^2 + e^{-2\beta J}\big)^{\frac{1}{2}}\Big) \tag{7.1.5}$$

as the elementary calculation of λ_+ shows, from (7.1.2). This is manifestly an analytic function of β, h in the region of physical interest (β, h real) so that the model has no phase transitions in the sense of no singularities of the thermodynamic functions or of their derivatives, as already discussed.

A similar method can be applied to the two-dimensional Ising model (Λ is now an $M \times N$ box). Suppose, for simplicity, $h = 0$, then $Z(\beta, h)$ is

7.1.6
$$\begin{aligned}&\sum_{\underline{\sigma}} \prod_{i=1}^{M} \prod_{j=1}^{N} e^{\beta J \sigma_{i,j}\sigma_{i+1,j} + \beta J \sigma_{i,j}\sigma_{i,j+1}} = \\ &= \sum_{\underline{\sigma}_1} \cdots \sum_{\underline{\sigma}_M} \prod_{i=1}^{M} \Big\{ \prod_{j=1}^{N} e^{\frac{\beta J}{2}\sigma_{i,j}\sigma_{i,j+1}} e^{\beta J \sigma_{i,j}\sigma_{i+1,j} + \frac{\beta J}{2}\sigma_{i+1,j}\sigma_{i+1,j+1}} \Big\}\end{aligned} \tag{7.1.6}$$

where in the second line we denote by $\underline{\sigma}_i = (\sigma_{i,1}, \ldots, \sigma_{i,N})$ all the spins on the i-th row of Λ; the periodic boundary conditions are imposed by setting $\underline{\sigma}_1 \equiv \underline{\sigma}_{M+1}$ and $\sigma_{i,1} \equiv \sigma_{i,N+1}$. Clearly, if we define the $2^N \times 2^N$ matrix

7.1.7
$$\begin{aligned}V_{\underline{\sigma},\underline{\sigma}'} &= \prod_{j=1}^{N} e^{\frac{\beta J}{2}\sigma_j\sigma_{j+1}} e^{\beta J \sigma_j \sigma'_j + \frac{\beta J}{2}\sigma'_j\sigma'_{j+1}} = \\ &= \exp \sum_{j=1}^{N} \big(\frac{\beta J}{2}\sigma_j\sigma_{j+1} + \beta J \sigma_j\sigma'_j + \frac{\beta J}{2}\sigma'_j\sigma'_{j+1} \big)\end{aligned} \tag{7.1.7}$$

where $\sigma_1 \equiv \sigma_{N+1}$, $\sigma'_1 = \sigma'_{N+1}$, we realize that

7.1.8
$$Z(\Lambda, \beta) = \operatorname{Tr} V^M \tag{7.1.8}$$

We have dealt so far only with periodic boundary conditions. We could introduce transfer matrices also in the case of other boundary conditions.

For instance, assume, for simplicity, that there are periodic boundary conditions along the columns; we shall consider the three cases below:

(1) "Perfect wall" or "open" boundary conditions, see §6.3, along the rows;

(2) Boundary conditions on the rows corresponding to the existence of fixed spins $\varepsilon_i = +1$ (or $\varepsilon_i = -1$) for all the i's on the lattice sites adjacent to the end points of the rows;

(3) Boundary conditions which are of the same type as in 2) but *half* the rows end in positive spins (say the upper half) and half in a negative spin.

We shall now write down a transfer matrix expression for $Z(\Lambda, \beta)$ in the above cases. In case (1) $Z(\Lambda, \beta) = \operatorname{Tr} V^{(1)M}$ where:

7.1.9
$$V^{(1)}_{\underline{\sigma},\underline{\sigma}'} = e^{\sum_{j=1}^{N-1} (\frac{\beta J}{2}(\sigma_j\sigma_{j+1} + \sigma'_j\sigma'_{j+1}) + \sum_{j=1}^{N} \beta J \sigma_j \sigma'_j)} . \tag{7.1.9}$$

In case (2) $Z(\Lambda,\beta) = \operatorname{Tr} V^{(\pm)M}$ where:

7.1.10
$$V^{(\pm)}_{\underline{\sigma},\underline{\sigma}'} = e^{\pm\beta J(\sigma_1+\sigma'_1+\sigma_N+\sigma'_N)} V^{(1)}_{\underline{\sigma},\underline{\sigma}'}\,. \tag{7.1.10}$$

In case (3), assuming here that the height of Λ is M with M even, we have that $Z(\Lambda,\beta) = \operatorname{Tr}(V^{(+)})^{\frac{1}{2}M-1}V^{(3+)}(V^{(-)})^{\frac{1}{2}M-1}V^{(3-)}$ with

7.1.11
$$V^{(3\pm)}_{\underline{\sigma},\underline{\sigma}'} = e^{\pm\beta J(\sigma'_1+\sigma'_N-\sigma_1-\sigma_N)} V^{(1)}_{\underline{\sigma},\underline{\sigma}'}\,. \tag{7.1.11}$$

The transfer matrix V in (7.1.7) is the matrix that was diagonalized in the famous paper of Onsager, [On44]. The matrix $V^{(1)}$ has also been diagonalized exactly in [Ab71].

The matrices $V^{(\pm)}$ have been studied and diagonalized in [AM73]. Many more exact calculations of interesting quantities have been performed, see for instance[Ab71], [MW73], [Ab78a], [Ab78b], [Ab84].

The problem of computing the partition function can be formulated similarly in the three dimensional case. Some very interesting results on the spectral properties of the generalization to three dimension of the matrix V (periodic boundary conditions) have been obtained in [MS70],[CF71].

In three dimensions one expects that the analogue of $V^{(3\pm)}$ (in contrast to $V^{(1)}$, $V^{(\pm)}$) has spectral properties which differ radically from those of V. In two dimensions the phenomenon should not occur and all the above matrices should have the same spectrum (asymptotically as $\Lambda \to \infty$). As mentioned in §6.15, problem (1), this should be related to the fact that $V^{(3\pm)}$ should contain some information about the rigidity of the line or surface of phase separation (which is "rigidly sitting" right near the two lines between which $V^{(3\pm)}$ "transfer", see Ch.VI).

A very interesting heuristic analysis of the spin correlation functions in terms of the transfer matrix has been done in [CF71].

§7.2. Meaning of Exact Solubility and the Two-Dimensional Ising Model

Before proceeding to study more interesting cases it is necessary to say that usually by "exactly soluble" one means that the free energy or some other thermodynamic function can be computed in terms of one or more quadratures, *i.e.* in terms of a finite-dimensional integral, whose dimension is *independent* of the system size.

In some cases one can even compute a few correlation functions: but there remain, as a rule, quite a few physically interesting quantities that one cannot compute (in the above sense of computing).

Another characteristic problem of the "solutions" is that sometimes their evaluation (in terms of quadratures) involve a few exchanges of limits that are not always easy to justify. For instance the value of the spontaneous magnetization in the Ising model was derived in an unknown way by Onsager (who just wrote the final formula, (7.2.2) below, on a blackboard at a

meeting in Firenze, [KO49]) and later by Yang, [Ya52], but both derivations relied on some hidden assumptions (in modern language the first assumed that there could be only two pure phases and the second that an external field of order $O(\frac{1}{N})$ would be strong enough to force the system into a pure magnetized phase).

In the case of the Ising model spontaneous magnetization the mathematical problems were solved much later, independently in [BGJS73] and [AM73]. But in several other cases there are still open problems in establishing the validity of the "exact solutions" with full rigor (in the current sense of the word).

The two-dimensional model is soluble only in zero external field ($h = 0$) in the sense that in this case the free energy $f_2(\beta)$, the magnetization $m(h)$ and a few correlation functions can be expressed quite simply. For instance if $\beta J^* \equiv J^*(\beta)$ is defined by $\tanh J^*(\beta) = e^{-2\beta J}$, see (6.6.6),(6.6.9), and denoting $\cosh^{-1}$ as the inverse function to cosh:

7.2.1
$$\begin{aligned}\beta f_2(\beta,0) = &\frac{1}{2}\log 2\sinh(2\beta J)+\\ &+\int_{-\pi}^{\pi}\frac{d\varphi}{4\pi}|\cosh^{-1}\big(\cosh 2\beta J\cosh 2J^* + \sinh 2\beta J\sinh 2J^*\cos\varphi\big)|\,.\end{aligned} \tag{7.2.1}$$

A simple analysis of the β dependence of this function shows that it is singular at the value $\beta = \beta_c$ for which it is $J = J^*$ (*i.e.* $\sinh 2\beta J = 1$) and the singularity appears as a logarithmic divergence of the derivative of f_2 with respect to β, *i.e.* as a divergence of the specific heat.

The exact value of the spontaneous magnetization, *i.e.* of the (right) h-derivatives at $h = 0$ of the free energy, can also be computed as said above and the result is

7.2.2
$$\lim_{h\to 0^+} -\frac{\partial\beta f_2}{\partial h}(\beta,h) = m(\beta) = \begin{cases}0 & \text{if } \sinh 2\beta J < 1\\ \big(1-(\sinh 2\beta J)^{-4}\big)^{\frac{1}{8}} & \text{otherwise}\,.\end{cases} \tag{7.2.2}$$

The importance of the above formulae, besides their obvious beauty, can hardly be overestimated. For instance they proved that in statistical mechanics there are phase transitions with critical exponents different from those of mean field theory: *e.g.* from (7.2.2) one realizes that $m(\beta) \xrightarrow[\beta\to\beta_c]{} 0$ as $(\beta-\beta_c)^{1/8}$ rather than the $(\beta-\beta_c)^{1/2}$ foreseen by mean field theory, see Ch.V, §5.1 and §5.2).

Many other quantities have been computed exactly: some already in the original Onsager papers and others in successive works. Among them we quote:

(1) The correlation function $\langle\sigma_{\underline{O}}\sigma_{\underline{x}}\rangle$ where $\underline{O}$ denotes the origin and $\underline{x}$ is a lattice point on one of the two lattice axes or, alternatively, it is a lattice point on the main diagonal; the symbol $\langle\cdot\rangle$ denotes the average value, of the quantity inside the brackets with respect to the Gibbs equilibrium distribution.

One shows that, if $\kappa(\beta)=2\beta|J-J^*|$ and if $|\underline{x}|$ is large, then the function $\langle\sigma_{\underline{O}}\sigma_{\underline{x}}\rangle^T \stackrel{def}{=} \langle\sigma_{\underline{O}}\sigma_{\underline{x}}\rangle-\langle\sigma_{\underline{O}}\rangle^2$ is proportional to

7.2.3
$$\langle\sigma_{\underline{O}}\sigma_{\underline{x}}\rangle^T \propto \begin{cases} \dfrac{e^{-\kappa(\beta)|\underline{x}|}}{\sqrt{|\underline{x}|}} & \text{for } \beta<\beta_c \\ |\underline{x}|^{-1/4} & \text{for } \beta=\beta_c \\ \dfrac{e^{-2\kappa(\beta)|\underline{x}|}}{|\underline{x}|^2} & \text{for } \beta>\beta_c \end{cases} . \tag{7.2.3}$$

One should not view (7.2.3) as a discontinuity in the asymptotic behavior of the pair correlation function when the temperature passes through the critical temperature T_c. A more detailed analysis shows in fact that if T is close to T_c the correlation starts depending on $\underline{x}$ *as if* $T=T_c$: this however proceeds only until $|\underline{x}|$ becomes so large to be comparable to the correlation length $\kappa(\beta)^{-1}$ which is longer and longer the closer T is to T_c. Afterwards the exponential decay sets in (with a different power correction depending on whether $\beta<\beta_c$ or $\beta>\beta_c$).

In fact one can compute the asymptotic behavior of all correlations functions (*i.e.* of the average value of products of spin values in an arbitrary number of sites) in various regimes: for instance for $\beta\neq\beta_c$ when the spin sites separate from each other homothetically, [WMTB76], [MTW77].

In the same situation a beautiful asymptotic formula for the value of the $2n$ spins correlation function for $2n$ spins aligned along a line has been derived by Kadanoff, [Ka69].

(2) The surface tension between coexisting phases, defined as, see (6.12.3), Ch.VI,

7.2.4
$$\tau(\beta)=\lim_{\Omega\to\infty}\frac{1}{L}\log\frac{Z^{+-}(\beta,\Omega)}{Z^{++}(\beta,\Omega)}=\begin{cases}0 & \text{if } \beta<\beta_c\\ 2\beta(J-J^*) & \text{if } \beta>\beta_c\end{cases} \tag{7.2.4}$$

where Z^{++}, Z^{+-} denote respectively the partition functions of the models obtained by fixing the boundary spins all equal to $+1$ in the first case and equal to $+1$ in the upper half and to -1 in the lower half in the second case. Here L is the perimeter of the container Ω, which is assumed to be a square. See [On44], [GMM72], [GM72b], [AM73].

(3) Quite a lot is known about correlation functions of spins associated with boundary sites. In the case of a box with open boundary conditions (*i.e.* no interaction with the spins located at external sites) it is, for instance, remarkable that for $\beta\to\beta_c$ the spontaneous magnetization on the boundary[1] does not tend to 0 as $\beta\to\beta_c$ with the same critical exponent $\frac{1}{8}$ characteristic of the bulk magnetization (*i.e.* of the magnetization at a site

---

[1] Whose square is defined here as the limit as $L\to\infty$ of the correlation $\langle\sigma_x\sigma_y\rangle$ with x,y being two points on $\partial\Lambda$ and at distance $O(L)$.

at a finite distance from the origin, hence infinitely far from the boundary): the new critical exponent is in fact $\frac{1}{2}$ (in the limit as $\Omega \to \infty$, of course), [Ab84].

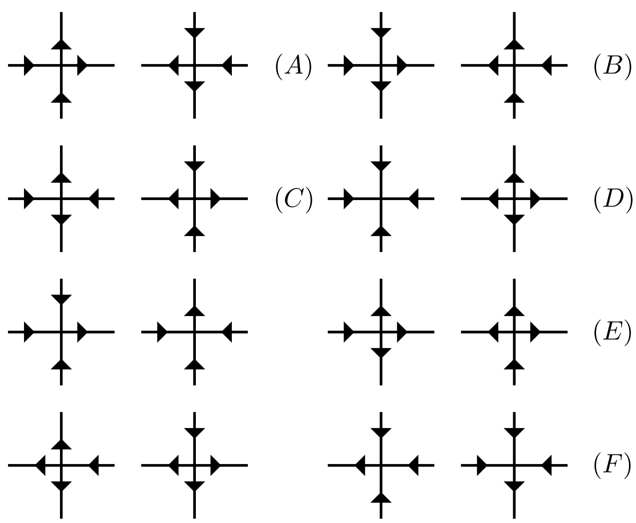
Even more remarkable is that one can find the critical exponent of the magnetization computed at a *corner*, where it is 1; more generally one considers the Ising model in a wedge-shaped planar lattice with opening angle ϑ (so that $\vartheta = \pi$ is the case of a point in the middle of a side of the container, whose shape does not matter as long as the point is at a distance tending to infitity from the points where the boundary starts bending: "half-plane case"); and $\vartheta = \frac{\pi}{2}$ is the case of a "square corner". The conjecture is that the critical exponent for the spontaneous magnetization at a "corner with opening ϑ" is $\frac{\pi}{2\vartheta}$. This has been proved (together with various other results about the magnetization at points a finite distance away from a corner) for $\vartheta = \pi$ (exponent 1, as mentioned above) and $\vartheta = \frac{\pi}{2}$ to which these corresponds an exponent $\frac{1}{2}$, see [AL95].

§7.3. Vertex Models.

Consider a rectangular region $\Omega \subset Z^2$ with opposite sites identified (periodic boundary conditions). We imagine that the microscopic states of the system are obtained by fixing an orientation on each lattice bond linking nearest neighbors of Ω.

Given a microscopic configuration $\underline{\sigma}$ of the system, at every lattice site we shall see one among the 16 possibilities shown in Fig. 7.3.1 below,

Fig. 7.3.1



The *eight-vertex models* or $8V$*-models* are characterized by allowing only the configurations $\underline{\sigma}$ which in *every lattice site* the bonds orientations look as in A, B, C, D, see p. 128, p. 203 in [Ba82]. Furthermore the energy associated with a configuration is, in the general eight-vertex model, a sum of contributions ε_j coming from each lattice site j. Allowing only vertices

A, B, C generates the *six-vertex models*, or $6V$*-models*.

The vertices A, B are called *polar vertices* while the C, D are called *non polar*: the A vertices are usually labeled as 1 and 2, the B vertices are labeled 3 and 4, the C are 5 and 6, the D are 7 and 8. We denote by c one of the latter eight arrow configurations: an $8V$ configuration will be a collection of arrows put on the lattice bonds that at each lattice node form one among the eight arrows configurations, called "allowed vertices".

Each lattice site x gives to the total energy of the configuration an additive contribution $\varepsilon(c)$ depending on which of the eight possible vertices c are formed by the arrows entering or exiting the lattice site x.

Not all the eight-vertex models are "exactly soluble": one gets a soluble model if the energies of the vertices $1, 2$, *i.e.* A, are equal and so are those of the vertices $3, 4$, *i.e.* B, of $5, 6$, *i.e.* C, and of $7, 8$, *i.e.* D. Thus the family of eight vertex soluble models has three parameters (because one of the four can be eliminated by recalling that the total energy is defined up to an additive constant).[2]

We call $\varepsilon_A \stackrel{def}{=} \varepsilon_1 = \varepsilon_2$ the common value of the energies corresponding to the vertices $1, 2$, and likewise we set $\varepsilon_B \stackrel{def}{=} \varepsilon_3 = \varepsilon_4$, $\varepsilon_C \stackrel{def}{=} \varepsilon_5 = \varepsilon_6$ and $\varepsilon_D \stackrel{def}{=} \varepsilon_7 = \varepsilon_8$.

The eight vertex models can be interpreted as spin models (hence as lattice gas models). A trivial way to do this is to interpret an arrow configuration as a spin configuration with the lattice of the spins being the lattice of the bonds, see Fig. 7.3.1; the up and down arrows can be identified with $+$ and $-$ spins located at the center of the arrows and, likewise, the right and left arrows can be identified with $+$ and $-$ spins. This naive procedure however relates the $8V$ models to spin models *with constraints* or *hard cores* because not all 16 configurations of spins on the arrows relative to a vertex are going to be possible.

A much more interesting representation of the eight vertex models is obtained by considering Ising models in which interactions between next nearest neighbor spins and *many spins* interactions occur (see §5.10) between quadruples of spins involving the four spins of a unit lattice cell, see p.207 in [Ba82]. An excellent introduction to the vertex models and their relationship with other models can be found in [LW72],[Ka74].

We call the lattice of the centers of the vertices the "$8V$ *lattice*" and we consider a configuration of arrows which at each point of the $8V$ lattice is one of the eight allowed configurations. We define a configuration of *signs* $+$ or $-$ located at midpoints of the bonds of the $8V$ lattice: $+$ represents a *up* or *right* arrow while $-$ represents a *down* or *left* arrow.

The product of the signs of the four bonds of the $8V$ lattice that merge into a vertex must be $+$: this is the condition that all vertices are of the above

---

[2] In fact the restriction $\varepsilon_5 = \varepsilon_6$ is not really such because the total energy depends only on $\varepsilon_5 + \varepsilon_6$. Furthermore if $\varepsilon_7 = \varepsilon_8 = 0$ the models with $\varepsilon_1 \neq \varepsilon_2$ and $\varepsilon_3 \neq \varepsilon_4$ are also soluble so that the class of soluble $6V$ models has four parameters, one more than the $8V$ model, see below.

first eight types. Therefore we have a *one-to-one* correspondence between the eight vertex configurations on the $8V$ lattice and the sign configurations at the centers of the bonds which multiply to $+$ on each of the $8V$ lattice bonds that merge into a vertex.

We imagine putting a *spin* $\sigma = \pm 1$ at the center of each square, also called a "*plaquette*", formed by four of the $8V$ lattice bonds; the lattice of the centers of the $8V$ lattice plaquettes will be called the *Ising lattice* and it will carry in this way a spin configuration $\underline{\sigma}$.

The *product* of the spins on nearest neighbor sites v, w of the Ising lattice is a *sign* ± 1 (we use this name rather than spin to stress the difference between these auxiliary variables and the spins introduced above). This sign can be naturally associated with the bond of the $8V$ lattice that separates v, w, and *by construction* the product of the signs on the sides that merge into a vertex is necessarily $+$ so that the sign configuration can be interpreted as an $8V$ configuration and there is a *one-to-two* correspondence between $8V$ configurations and Ising spin configurations. In fact a spin configuration $\underline{\sigma}$ and $-\underline{\sigma}$ give the same sign configuration, hence the same $8V$ configuration.

Setting $a = \exp(-\beta\varepsilon_A), b = \exp(-\beta\varepsilon_B), c = \exp(-\beta\varepsilon_C), d = \exp(-\beta\varepsilon_D)$ and defining J, J', J'' :

7.3.1
$$\begin{array}{ll} a = \exp\beta(J + J' + J'') & b = \exp\beta(-J - J' + J'') \\ c = \exp\beta(-J + J' - J'') & d = \exp\beta(J - J' - J'') \end{array} \tag{7.3.1}$$

it is now immediate to check that any configuration of the $8V$ model has the same energy as the corresponding spins configuration $\pm\underline{\sigma}$ in the Ising model with energy:

7.3.2
$$H(\underline{\sigma}) = -\Big(\sum J\sigma_i\sigma_{i'} + \sum J'\sigma_i\sigma_{i''} + \sum J''\sigma_i\sigma_j\sigma_{j'}\sigma_{j''}\Big) \tag{7.3.2}$$

where the sum runs over the sites $i \in \Omega$ and i' denotes the nearest neighbor of i along the diagonal of the second and fourth quadrant, and i'' the one along the first and third quadrant diagonal; j, j', j'' are three sites that, together with i, form a unit square (with i in the lower left corner), see p. 207 in [Ba82].

One can consider also the sixteen-vertex models, obtained by considering *all* arrow configurations in Fig. 7.3.1 above, including the E,F ones. This model is also equivalent to a suitable Ising model, with also three-spin interactions, see [LW72], p. 350, for the discussion of the general cases.

This model has many interesting special cases, some of which were recognized to be soluble before Baxter's work. In fact the breakthrough in the whole theory was the solution by Lieb of the six-vertex Pauling's "Ice model" discovering the method of solution for the other soluble $6V$-models.

Among the latter there are the *six-vertex models*, whose configurations only allow for the vertices A, B, C each of which gives a contribution to the energy $\varepsilon_A, \varepsilon_B, \varepsilon_C$.

(1) The just mentioned Pauling's *ice model* fixes $\varepsilon_A = \varepsilon_B = \varepsilon_C = 0$, (it corresponds to $J' = J_0, J = -J_0, J'' = J_0$ and $J_0 = +\infty$).

(2) The KDP model fixes the B, C vertices energies to $\varepsilon > 0$ and to 0 those of the vertices A: the occurrence of "non polar" vertices and of two of the four polar ones is energetically less favored.

(3) The F model fixes the energies of A, B to $\varepsilon > 0$ and those of C to 0: the non polar vertices are favored.

Such models were solved (*i.e.* their free energy was calculated) in rapid succession after the solution of the first, [Li67a], [Li67b], [Li67c], [Su67]. The analyis and solution of the most general 6V model is in [LW72].

The $6V$ models above are limiting cases of the $8V$ model in which the couplings J, J', J'' in (7.3.2) tend (suitably) to ∞. Because of this limiting procedure they have been sometimes regarded as "pathological", see below.

The $6V$ models have a physical meaning within the theory of the hydrogen bond and of similar chemical bonds. In the ice model the lattice sites represent the sites occupied by O, and the bond orientations tell where the two H atoms are located: if an arrow emerges from a lattice site this means that an H atom is located on the bond and near the site. The association of the arrows with the bond provides a two-dimensional version of the *ice rule* (which states that on any bond there is one H atom, and not more, located closer to one of the two oxygens), a rule deduced by Pauling from the observation that the ice entropy is lower than the one it would have if in an ice crystal the H atoms could be found, unconstrained, near every O atom (so that one could find configurations like E, F or even with opposite arrows on each bond). Of course the model should be, to be realistic, three dimensional, but the appropriate three-dimensional version is not exactly soluble.

The KDP-model has been proposed as a model for the ferroelectric properties of KH_2PO_4: a substance that crystallizes in tetrahedra with KPO_4 at the center and the two H atoms on the lines between the KPO_4: only one H can be located on each such line and it can be there in two positions (*i.e.* near one or the other extreme). KH_2PO_4 is a polar molecule without spherical symmetry so that not all dipoles give equal contribution to the total energy of a configuration. In the two-dimensional version of the model the two nonpolar vertices C and two of the polar ones (*e.g.* B) are unfavored and contribute energy $\varepsilon > 0$ while the others contribute $\varepsilon = 0$: at low temperature a *spontaneous polarization*, or *ferroelectricity*, is expected to occur.

The F-model, instead, is a model for an *antiferroelectric* polar material resisting (at low temperature at least) polarization by a field.

A deeper discussion of the physical interpretation of the 6V models and of their relation with other remarkable combinatorial problems and statistical mechanics models can be found in [LW72]. The vertex models are equivalent in various senses to several other models, see for instance, [VB77], [Ba82].

The problem of the existence of the thermodynamic limit for the 8V models is a special case of the general theory, see Ch.IV, with the minor modifica-

tion required to study lattice spin systems rather than continuous particle systems (see (7.3.2): an easier problem, in fact). *However the situation is different in the case of the 6V models*, because the constraints imposed by arrow configurations can propagate the boundary conditions on a large box all the way inside it. The theory has been satisfactorily developed in [LW72].

In the ice model case one finds, [Li67a]

7.3.3
$$\lim_{\Omega\to\infty}\frac{1}{|\Omega|}\log Z(\Omega)=\frac{3}{2}\log\frac{4}{3} \tag{7.3.3}$$

which is really beautiful!

In the F-model case the free energy is given by, if $\Delta = 1-\frac{1}{2}e^{2\beta\varepsilon}$,

7.3.4
$$\begin{aligned}&-\beta f_F(\beta)=-\beta\varepsilon+\\&+\begin{cases}\frac{1}{8\mu}\int_{-\infty}^{\infty}\log\frac{\cosh\alpha-\cos 2\mu}{\cosh\alpha-1}\frac{d\alpha}{\cosh\pi\alpha/\mu} & \text{if } \cos\mu=|\Delta|<1\\ \frac{\lambda}{2}+\sum_{n=1}^{\infty}\frac{e^{-n\lambda}\tanh n\lambda}{n} & \text{if } \cosh\lambda=-\Delta>1\end{cases}\end{aligned} \tag{7.3.4}$$

and in the KDP-model case, setting $\Delta=\frac{1}{2}e^{\beta\varepsilon}$, the free energy is:

7.3.5
$$\begin{aligned}&-\beta f_{KDP}(\beta)=\\&=\begin{cases}\frac{1}{8\mu}\int_{-\infty}^{\infty}\frac{\cosh\alpha-\cos\mu}{\cosh\alpha-\cos 3\mu}\frac{d\alpha}{\cosh\pi\alpha/2\mu} & \text{if } \Delta=-\cos\mu<1\\ 0 & \text{otherwise}\end{cases}\end{aligned} \tag{7.3.5}$$

where the above solutions for the KDP-model is due to Lieb, [Li67b], and for the F-model to Lieb and Sutherland, [Li67c], [Su67]. See §7.5 below for a technical introduction.

Furthermore the F-model and the KDP-model, unlike the soluble eight vertex models, can be solved even in the presence of an "electric field" E, if such field is modeled by assuming that the energy contribution of a vertex increases by $-Ep$ with p being the number of arrows pointing up minus the number of those pointing down. The solution in the presence of such an electric field is fairly simple but we do not report it here. It is a model with one more free parameter in which the energies $\varepsilon_1,\varepsilon_2$ are different and the energies $\varepsilon_3,\varepsilon_4$ are also different by the same amount.

The even more general model in which the electric field has also a horizontal component so that the energy of a vertex increases also by $-E'q$ with q being the number of arrows pointing right minus the number of those pointing left *is also soluble* although the result is not as simple to discuss, [LW72]. Taking this into account amounts to a freedom of taking ε_j, $j=1,\ldots,4$ as independent parameters. Hence the number of free parameters in the most general $6V$ soluble model is 4 (they are $\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4,\varepsilon_5=\varepsilon_6$ and one can be fixed to be 0).

The elementary analysis of the above formulae, and of their extensions to cases with $E\neq 0$, leads to the following results that we describe by

denoting by $f_F(\beta, E)$ and $f_{KDP}(\beta, E)$ the free energies of the two models at temperature $(k_B\beta)^{-1}$ and in the presence of a (vertical) electric field E.

The function $f_F(\beta, 0)$ is infinitely differentiable and analytic in β except for an essential singularity, at the value $\beta = \beta_c$ where $\Delta = -1$; one has, therefore, in zero field a phase transition of *infinite order* (the order of a transition as a function of a parameter can be defined to be the order of the lowest derivative of the free energy, with respect to the parameter, which is singular, see Ch.VI, §6.5).

The polarization, defined as the average number of arrows pointing up minus those pointing down, is proportional to the derivative of f_F with respect to E; it vanishes for $E = 0$ for all values of β; but if $\beta > \beta_c$ (low temperature) it remains 0 *even* if $E \neq 0$, for a while, and it becomes nonzero only if E grows beyond a critical value $E_c(\beta)$ and in this sense the model has an *antiferromagnetic* behavior. If one keeps E fixed, then by varying β one finds a *second-order* phase transition with a specific heat singularity proportional to $(\beta - \beta_c)^{-1/2}$.

The free energy of the KDP-model, $f_{KDP}(\beta, E)$, is essentially different. Also in this case there is a critical temperature $\beta = \beta_c$ at zero field E (defined by $\Delta = 1$): in zero field and if $\beta > \beta_c$ the polarization has value 1 identically, and the free energy is constant; if $\beta \to \beta_c^-$ the specific heat tends to 0 as $(\beta_c - \beta)^{1/2}$ but the internal energy does not tend to zero although the value of the internal energy for $\beta > \beta_c$ is 0, therefore there is a phase transition of first order with latent heat and at low temperature there is spontaneous polarization (maximal, $p = 1$, so that the system is "frozen" and it has trivial thermodynamic functions).

The above properties, selected among many that can be derived by simply examining the expressions for the exact solutions, show the richness of the phenomenology and their interest for the theory of phase transitions, in particular as examples of phase transitions with properties deeply different from those found in the Ising model case, [LW72].

It is important to stress that the $6V$ models form a four-parameters class of models and some of the soluble $6V$ models are limiting cases of the three-parameter family of soluble $8V$ models. By varying the parameters one can find a continuous path linking the F-model critical point (in zero field) to the KDP-model critical point (also in zero field).

As remarked above the $8V$ models are genuine short-range Ising models, (7.3.2), with *finite couplings*. Hence one can study how the critical point singularity changes in passing from the F to the KDP-model. The remarkable result found via the solution of the $8V$ models is that the critical exponents (the ones that can be computed) change *continuously* from the F values to the KDP values.

This fact has great importance: at the time of the solution of the ice model and of the consequent solution of the F and KDP models the universality theory of critical point singularity was not yet developed in its final form. So when the renormalization group approach arose around 1969, see [BG95] for a review and references, the $6V$ model appeared as a counterexample to

the universality that the renormalization group was supposed to predict.

This was stressed by Lieb on several occasions: regrettably his comments were either not understood or they were dismissed on the grounds that the $6V$ models regarded as Ising models were "*models with constraints*", in the sense discussed above in the derivation of (7.3.2).

Baxter's solution of the $8V$ model made clear to everybody Lieb's point of view. Even genuine finite-range "unconstrained" Ising models could show a critical singularity that was neither that of the F-model, neither that of the KDP-model nor that of the Ising model (to which the $8V$ models reduce if $J'' = 0$). The three behaviors were the limiting cases of a three-parameter continuum of possibilities! so that the phenomena shown by the $6V$ models were not examples of pathological properties dubiously attributable to the hard core features of such models,[3] but they were the rule! This led to a much better understanding of the theories that were put forward to explain the universality phenomena, first among which the renormalization group itself.

It would take too long to discuss the eight-vertex model properties, [Ba82]: it is not surprising that it offers a varied and interesting phenomenology, besides the enormous theoretical interest of the sophisticated analysis necessary to obtain the solutions. As mentioned it can be solved only in zero field (*i.e.* for $\varepsilon_1 = \varepsilon_2$, $\varepsilon_3 = \varepsilon_4$, $\varepsilon_5 = \varepsilon_6$, $\varepsilon_7 = \varepsilon_8$). Some results in nonzero field can be obtained by "perturbing" the zero field models, see [MW86].

It is however important to stress once more that the exactly soluble models give very limited information about the actual thermodynamics of the system. For instance the informations that can be obtained about correlations, even just pair correlation functions, is very scanty.

There are a few remarkable cases in which one can compute explicitly the pair correlation function, like the two-dimensional Ising model, see [MW73], [WMTB76], [MTW77], [AR76]), or even higher correlations, like the $2n$-spins correlations in the two-dimensional Ising model when the spins are on the same lattice line and far apart, [Ka69]. Correlation inequalities can be very useful to study more general situations without having recourse to exact solutions, [MM77].

Often one is able to evaluate the *correlation length* because it can be related to the second or third largest eigenvalue of the transfer matrix (often the highest is almost degenerate and what counts is the third) which can be studied quite explicitly (for instance in the $8V$ models), see [Ba82] p. 241,284.

Even the latter results provide little information about the "critical" cases when there is no gap isolating the top of the transfer matrix spectrum from the rest. The renormalizations group methods, which received so much clarification from the exact solution of the $6V$ and $8V$ models, do provide at least in some regions of the coupling parameters a, b, c, d (7.3.1), a rather

[3] Because as mentioned above their statistical mechanics properties are quite normal, as shown in [LW72], p. 354-361.

detailed analysis of the properties of the correlation decay. For a recent development see [Ma98].

There are many other exactly soluble models; *e.g.* one could just mention the *spherical model*, [BK52], the *dimer model*, [Ka61] and [TF61], the *XY-model*, the ground state of the one-dimensional Heisenberg model (whose solution can be related to that of the six and eight vertex models [YY66],[Ba82], see also [Li67a],[Su70],[Ka74]), besides the Ising model on some lattices other than the square lattice and the *hard hexagon model* on a triangular lattice, [Ba82]. The reader should consult the monographs on the subject [MW73], [Ba82] to which a good introduction is still the review paper [SML64] and the book [LM66].

§7.4. A Nontrivial Example of Exact Solution: the Two-Dimensional Ising Model

The actual computations for the exact solutions are always quite involved but their elegance surpasses Jacobi's theory of the action angle variables for the pendulum. To give an idea of the procedures involved we describe below the classical solution of the Ising model, as presented in the paper [SML64].[4]

One starts from the remark in §6.15 that the free energy of the Ising model in an $N\times M$ box with periodic boundary conditions is given by the trace (7.1.8) of the M-th power of the matrix V in (7.1.7).

It is easy to construct a convenient representation for the matrix V. Consider the three Pauli matrices

7.4.1
$$\sigma^x=\begin{pmatrix}0&1\\1&0\end{pmatrix},\qquad \sigma^y=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\qquad \sigma^z=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\tag{7.4.1}$$

and consider the tensor product $\mathcal{H}$ of N bidimensional linear spaces E and the operators σ_j^α defined by:

7.4.2
$$\begin{aligned}\mathcal{H}=&\prod_{j=1}^N E=E\otimes E\otimes\ldots\otimes E\\ \sigma_j^\alpha=&I\otimes\ldots\otimes\sigma^\alpha\otimes I\otimes\ldots\otimes I,\qquad \alpha=x,y,z\end{aligned}\tag{7.4.2}$$

where I is the identity operator on E and σ^α is located at the jth-place.

The operators σ_j^z are pairwise commuting and one can easily diagonalize them. If $|\sigma\,\rangle\in E$ is a vector such that $\sigma^z\,|\sigma\,\rangle=\sigma\,|\sigma\,\rangle$ (with $\sigma=\pm1$), then the most general eigenvector of the operators σ_k^z on $\mathcal{H}$ can be written as a tensor product of vectors:

---

[4] This was important progress as it provided a simple and easily understandable entirely new approach to the solution of the Ising model, at a time when rested on the original works [On44], [Ka49], [Ya52] which were still considered very hard to follow in the 1960s.

7.4.3
$$|\underline{\sigma}\rangle = \prod_{j=1}^{N}{}^{\otimes} |\sigma_j\rangle\,, \qquad \underline{\sigma} = (\sigma_1, \ldots, \sigma_N) \tag{7.4.3}$$

and the 2^N such vectors (each corresponding to a string $\underline{\sigma} = (\sigma_1, \ldots, \sigma_N)$ of digits $\sigma_j = \pm 1$) form an orthonormal basis in $\mathcal{H}$.

Consider the operator w on E such that

7.4.4
$$\langle\, \sigma \,|\, w \,|\sigma'\,\rangle \equiv e^{\beta J \sigma\sigma'} \qquad \sigma, \sigma' = \pm 1 \tag{7.4.4}$$

which can be written as a matrix as

7.4.5
$$w = \begin{pmatrix} e^{\beta J} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta J} \end{pmatrix} = e^{\beta J} I + e^{-\beta J}\sigma^x = A e^{J^* \sigma^x} \tag{7.4.5}$$

with $A = (2\sinh 2\beta J)^{1/2}$ and $\tanh J^* = e^{-2\beta J}$. We see from (7.4.4) that the tensor product

7.4.6
$$V_1 = \prod_{j=1}^{N}{}^{\otimes} w = w \otimes \ldots \otimes w = A^N e^{J^* \sum_j \sigma_j^x} \tag{7.4.6}$$

has matrix elements

7.4.7
$$\langle\, \underline{\sigma} \,|\, V_1 \,|\underline{\sigma}'\,\rangle = \prod_{j=1}^{N} e^{\beta J \sigma_j \sigma_j'}\,. \tag{7.4.7}$$

Hence if we define the operator V_2 on $\mathcal{H}$, diagonal on the basis consisting in the vectors $|\underline{\sigma}\rangle$, which on $|\underline{\sigma}\rangle$ acts as

7.4.8
$$V_2 = e^{\frac{1}{2}\beta J \sum_j \sigma_j^z \sigma_{j+1}^z} \tag{7.4.8}$$

then we immediately check that the matrix elements of the operator $V = V_2 V_1 V_2$ between $|\underline{\sigma}'\rangle$ and $\langle\,\underline{\sigma}\,|$ are exactly the transfer matrix elements $V_{\underline{\sigma},\underline{\sigma}'}$ in (7.1.7).

This means that the problem of the evaluation of the partition function (7.1.8) is solved once we know the eigenvalues of $V = V_2 V_1 V_2$, or of any operator unitarily equivalent to it.

We shall, therefore, perform the unitary transformation U on $\mathcal{H}$ such that:

7.4.9
$$U\sigma_j^x U^{-1} = \sigma_j^z \qquad U\sigma_j^z U^{-1} = -\sigma_j^x \tag{7.4.9}$$

which transforms the matrix V into $\tilde{V} = UVU^{-1}$:

7.4.10
$$\tilde{V} = A^N \Big(e^{\frac{1}{2}\beta J \sum_j \sigma_j^x \sigma_{j+1}^x} \Big) \Big(e^{J^* \sum_j \sigma_j^z} \Big) \Big(e^{\frac{1}{2}\beta J \sum_j \sigma_j^x \sigma_{j+1}^x} \Big) \tag{7.4.10}$$

which we can write, defining the matrices $\tilde{V}_j$, as $\tilde{V}_2 \tilde{V}_1 \tilde{V}_2$.

Here it will be very usefule to consider the following *Pauli-Jordan transformation*:

7.4.11
$$\sigma_j^{\pm} = \frac{1}{2}(\sigma_j^x \pm i\sigma_j^y), \qquad a_j^{\pm} = \sigma_j^{\pm} \prod_{s=1}^{j-1}(-\sigma_s^z) \tag{7.4.11}$$

so that $\sigma_j^{\pm} = a_j^{\pm} \prod_{s<j}(1 - 2a_s^+ a_s^-)$ and $\sigma_j^+\sigma_j^- = a_j^+ a_j^-$. The usefulness of (7.4.11) is due to the remark that:

7.4.12
$$[a_j^+, a_{j'}^+]_+ = 0 = [a_j^-, a_{j'}^-]_+, \qquad [a_j^+, a_{j'}^-]_+ = \delta_{jj'} \tag{7.4.12}$$

if $[\cdot]_+$ denotes the *anticommutator*. In other words the above transformations changes the operators $\sigma_j^{\pm}$ which have *mixed commutation relations* (*i.e.* commuting for $j \neq j'$ while anticommuting for $j = j'$) into operators $a_j^{\pm}$ with purely *fermionic commutations relations*. We set $\mathcal{N} = \sum_j a_j^+ a_j^-$.

This makes it easy to complete the calculations: one first remarks that the *parity operator* $(-1)^{\mathcal{N}} \stackrel{def}{=} (-1)^{\sum_j \sigma_j^+\sigma_j^-} \equiv (-1)^{\sum_j a_j^+ a_j^-}$ commutes with the factors defining $\tilde{V}$, hence it commutes with $\tilde{V}$ itself.

Hence the space $\mathcal{H}$ is a direct sum of two orthogonal subspaces $\mathcal{H}^+$ and $\mathcal{H}^-$ on which the parity operator has values respectively $+1$ and -1, *i.e.* $\mathcal{H}^+$ contains an "even number of fermions" and $\mathcal{H}^-$ an "odd number".

This means that if $|\Omega\rangle$ is a vector in $\mathcal{H}$ such that $a_j^- |\Omega\rangle \equiv 0$ for $j = 1, \ldots, N$, and one can see that there is always one and only one such vector, the $\mathcal{H}^+$ is the space spanned by the vectors $a_{j_1}^+ \ldots a_{j_{2n}}^+ |\Omega\rangle$, $n \geq 0$ while $\mathcal{H}^-$ is spanned by the vectors with $a_{j_1}^+ \ldots a_{j_{2n+1}}^+ |\Omega\rangle$, $n \geq 0$.

Assuming from now on that $\frac{1}{2}N$ is *even*, for simplicity, one can make the following key remark; if we define the new operators:

7.4.13
$$A_q^{\pm} = \frac{e^{\pm i\pi/4}}{\sqrt{N}} \sum_{j=1}^{N} e^{\pm ijq} a_j^{\pm}, \qquad q = \pm\frac{\pi}{N}, \pm 3\frac{\pi}{N}, \ldots \pm (N-1)\frac{\pi}{N} \tag{7.4.13}$$

where the q's are defined so that $e^{iqN} = -1$, then *on* $\mathcal{H}^+$ the following algebraic identities hold:

7.4.14
$$\begin{aligned} a_j^{\pm} =& \frac{e^{\mp i\pi/4}}{\sqrt{N}} \sum_q e^{\mp ijq} A_q^{\pm} \\ \tilde{V}_1 =& A^N e^{J^* \sum_q (2A_q^+ A_q^- - 1)} \\ \tilde{V}_2 =& e^{\frac{1}{2}\beta J \sum_q \left((A_q^+ A_q^- + A_{-q}^+ A_{-q}^-) \cos q + (A_q^+ A_{-q}^+ + A_{-q}^- A_q^-) \sin q \right)} \end{aligned} \tag{7.4.14}$$

where, in deriving the second and third relations, careful account has been taken of the facts that while $\sigma_j^x \sigma_{j+1}^x = (\sigma_j^+ + \sigma_j^-)(\sigma_{j+1}^+ + \sigma_{j+1}^- = (a_j^+ - a_j^-)(a_{j+1}^+ + a_{j+1}^-)$ for $j < N$ it is, *instead*, $(\sigma_N^+ + \sigma_N^-)(\sigma_1^+ + \sigma_1^-) = -(-1)^{\mathcal{N}}(a_N^+ - a_N^-)(a_1^+ + a_1^-)$ and that on $\mathcal{H}^+$ it is $(-1)^{\mathcal{N}} = +1$.

The anticommutation relations (7.4.12) imply similar anticommutation relations for the $A^{\pm}$ so that the expressions in (7.4.14) containing different $\pm q$'s *commute* and therefore we can write, *only on*, $\mathcal{H}^+$,

7.4.15
$$\tilde{V} = A^N \prod_{q>0} P_q \tag{7.4.15}$$

with A defined after (7.4.5) and

7.4.16
$$P_q = e^{\beta J\left((A_q^+ A_q^- + A_{-q}^+ A_{-q}^-)\cos q + (A_q^+ A_{-q}^+ + A_{-q}^- A_q^-)\sin q\right)} . \tag{7.4.16}$$
$$e^{2J^*(A_q^+ A_q^- + A_{-q}^+ A_{-q}^- - 1)} \cdot e^{\beta J\left((A_q^+ A_q^- + A_{-q}^+ A_{-q}^-)\cos q + (A_q^+ A_{-q}^+ + A_{-q}^- A_q^-)\sin q\right)}$$

and the operators P_q commute. *Hence* they can be considered as 4×4 matrices acting on the space D_q spanned by the vectors $|\Omega\rangle$, $A_q^+|\Omega\rangle$, $A_{-q}^+|\Omega\rangle$, $A_q^+ A_{-q}^+|\Omega\rangle$.

Furthermore it is clear that the operator P_q has $A_q^+|\Omega\rangle$ as eigenvectors and $A_{-q}^+|\Omega\rangle$ with eigenvalues $e^{2\beta J\cos q}$. Hence the $\frac{N}{4}$ operators P_q can be considered as operators on the space D_q^+ spanned by the two vectors $|\Omega\rangle, A_q^+ A_{-q}^+|\Omega\rangle$. Note that such space is invariant under the action of P_q.

Diagonalization of (7.4.15) will account only for $\frac{1}{2}2^N$ eigenvalues and eigenvectors because the operator $\tilde{V}$ coincides with the transfer matrix *only oon* $\mathcal{H}^+$. The remaining $\frac{1}{2}2^N$ are in the space $\mathcal{H}^-$ and they can be found in the same way: defining new operators $A_q^{\pm}$ as in (7.4.13) with q such that $e^{iqN} = 1$ one obtains a representation like (7.4.15) which is now only correct if restricted to the space $\mathcal{H}^-$.

Therefore the problem has been reduced to the diagonalization of $\frac{N}{4}$ 2×2 matrices obtained by restricting each P_q, $q > 0$, to the space spanned by the two vectors $A_q^+ A_{-q}^+|\Omega\rangle, |\Omega\rangle$. The four matrix elements of P_q on such vectors can be evaluated easlily starting from the remark that

7.4.17
$$P_q = e^{2\beta J\cos q} e^{\beta J(\tau^z\cos q + \tau^x \sin q)} e^{2J^*\tau^z} e^{\beta J(\tau^z\cos q + \tau^x\sin q)} \tag{7.4.17}$$

where

7.4.18
$$\begin{aligned} \tau^z =& (A_q^+ A_q^- + A_{-q}^+ A_{-q}^- - 1), \qquad \tau^x = (A_q^+ A_{-q}^+ + A_{-q}^- A_q^-) \\ \tau^y =& \frac{1}{i}(A_q^+ A_{-q}^+ - A_{-q}^- A_q^-) \end{aligned} \tag{7.4.18}$$

and the matrices τ^x, τ^y, τ^z have the same matrix elements as $\sigma^x, \sigma^y, \sigma^z$ in (7.4.1), respectively, if the vectors $A_q^+ A_{-q}^+|\Omega\rangle, |\Omega\rangle$ are identified with $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$, hence they satisfy the same commutation and multiplication relations and have the same spectrum.

By using the properties of the Pauli matrices it is possible, with some obvious algebra, to rewrite the product (7.4.17) as

7.4.19
$$P_q = e^{2\beta J\cos q}\left(\cosh\varepsilon(q) + (\tau^z\cos\vartheta_q + \tau^x\sin\vartheta_q)\sinh\varepsilon(q)\right) \qquad (7.4.19)$$

where $\varepsilon(q)$ is the positive solution of

7.4.20
$$\cosh\varepsilon(q) = (\cosh 2J^*\,\cosh 2\beta J + \sinh 2J^*\,\sinh 2\beta J\,\cos q) \qquad (7.4.20)$$

and the angle ϑ_q is defined, setting $t = \tanh\beta J$ and $t^* = \tanh J^*$, by

7.4.21
$$e^{2i\vartheta_q} = \frac{(1 + e^{iq}\frac{t}{t^*})(1 + tt^*e^{iq})}{(1 + e^{-iq}\frac{t}{t^*})(1 + tt^*e^{-iq})} \qquad (7.4.21)$$

Coming back to P_q one sees that on D_q^+:

7.4.22
$$P_q = e^{2\beta J\cos q}e^{\varepsilon(q)(\tau^z\cos\vartheta_q + \tau^x\sin\vartheta_q)} \qquad (7.4.22)$$

but since the latter expression has $A_q^+\,|\Omega\rangle$ and $A_{-q}^+\,|\Omega\rangle$ as eigenvectors with the "correct" eigenvalues $e^{2\beta J\cos q}$ *we see that P_q is given by (7.4.22) on the entire four-dimensional space D_q*. Hence we must diagonalize:

7.4.23
$$\tilde{P}_q = e^{2\beta J\cos q}e^{\varepsilon(q)\left((A_q^+A_q^- + A_{-q}^+A_{-q}^- - 1)\cos\vartheta_q + (A_q^+A_{-q}^- + A_{-q}^-A_q^-)\sin\vartheta_q\right)}. \qquad (7.4.23)$$

Note that if one sets (for $\varphi_q + 2\vartheta_q = 0$)

7.4.24
$$\begin{aligned} A_q^+ &= B_q^+\cos\varphi_q + B_{-q}^-\sin\varphi_q \\ A_{-q}^- &= -\,B_q^+\sin\varphi_q + B_{-q}^-\cos\varphi_q \end{aligned} \qquad (7.4.24)$$

which is an example of a well-known transformation, called the "Bogoliubov-Valatin transformation", and $\varphi_q = -2\vartheta_q$ one finds

7.4.25
$$\tilde{P}_q = e^{2\beta J\cos q}e^{\varepsilon(q)(B_q^+B_q^- + B_{-q}^+B_{-q}^- - 1)} \qquad (7.4.25)$$

so that the transfer matrix on $\mathcal{H}^+$ can be written as $A^N\exp(\sum_q(\beta J\cos q + \varepsilon(q)(B_q^+B_q^- - \frac{1}{2}))$, see (7.4.15), noting that $\varepsilon(q) = \varepsilon(-q)$ the $\frac{1}{2}2^N$ eigenvalues of the transfer matrix on $\mathcal{H}^+$ are, therefore, simply given by $A^N e^{\sum_q \frac{1}{2}e_q}$, $e_q = \pm\varepsilon(q)$ where the signs can be arbitrarily chosen as long as the is an even number of + signs (because the only relevant eigenvalues are those associated with an even number of fermions). The terms $\beta J\cos q$ disappear because of the trigonometric identity $\sum_q \cos q = 0$.

There is a unique vector $|\Omega'\rangle$ such that $B_q^-\,|\Omega'\rangle \equiv 0$ for all $q > 0$, and it is a linear combination of the vectors obtained by applying $\prod_k A_{q_k}^+A_{q_k}^-$ to $|\Omega\rangle$: this means that the "new vacuum" $|\Omega'\rangle$ is in the even space $\mathcal{H}^+$.

Clearly the largest eigenvalue is $A^N \sum_{q>0} \frac{1}{2}\varepsilon(q)$ and it corresponds to the eigenvector $\prod_q B_q^+ |\Omega'\rangle$ which, being in $\mathcal{H}^+$, really corresponds to an eigenvector of the transfer matrix.

A similar calculation can be performed to find the eigenvalues of the transfer matrix on $\mathcal{H}^-$: this time one takes q such that $e^{iqN} \equiv 1$ and proceeds in the same way getting an analogous result. The analysis of the relation between the maximal eigenvalues in $\mathcal{H}^+$ and $\mathcal{H}^-$ is a simple matter: however it is very illuminating as it reveals that if $\beta < \beta_c$, with β_c defined by $\beta_c J = J^*(\beta_c)$, there is a gap between the two eigenvalues (which does not shrink to 0 as $N \to \infty$) while for $\beta > \beta_c$ (*i.e.* at low temperature) the difference between the two eigenvalues tends to 0 exponentially fast as $N \to \infty$. In the end this can be seen as the reason for the different asymptotic behavior of the correlations, in (7.2.3)).

In the analysis on $\mathcal{H}^-$ one must pay attention to the fact that now the values $q = 0, \pi$ are permitted but they are not paired as the other values of q. This implies that the modes $q = 0, \pi$ must be treated differently, and the key difference that a careful discussion yields is that the value of $e(\pi)$ is *not necessarily* > 0: it becomes negative for $T < T_c$ and this accounts for the difference in the spectrum of the transfer matrix above and below T_c.

With the above explicit expressions for the eigenvalues it is not difficult to see that in the limit as $N \to \infty$ the only eigenvalues that count are the two with maximum modulus, which are almost degenerate if $\beta > \beta_c$ and separated by a gap of order $O(1)$ for $\beta < \beta_c$. This means that the limit as $N \to \infty$ of (7.1.8) is dominated by the largest eigenvalue λ_+ and the free energy $\beta f(\beta)$ is always (for all β) given by $\log A$ plus the limit of $\frac{1}{N}\sum_{q>0} \varepsilon(q)$ which is the integral in (7.2.1) as a consequence of (7.4.20).

The calculation of the spontaneous magnetization is more involved (see [Ya52], it is, however, quite simple in the method of [SML64]), and it requires extra assumptions, which can be removed by using further arguments as mentioned in Ch.VI, [BGJS72], [AM73].

We see that the above calculation requires some wit but its real difficulty is to realize that the transfer matrix can be written essentially as *a quadratic form in certain fermionic operators*. This reduces the problem to a 4×4 matrix diagonalization problem; after that it is clear that the problem is "solved", although some computations are still necessary to get a really explicit expression for the free energy.

The above calculation should be performed in all its details by those interested in statistical mechanics, as it is one of the high points of the theory, in spite of its apparent technicality. There are alternative ways to compute the free energy of the two-dimensional Ising model, but all of them have a key idea and a lot of obvious technicalities that accompany it. The one above is particularly interesting because of its connection with the quantum theory of fermionic systems (and because of its simplicity).

§7.5. The Six Vertex Model and Bethe's Ansatz

We consider an $N \times M$ periodic lattice (N, M even) and on each bond we put an arrow so that the arrows entering or leaving each lattice node form a configuration c among the A,B,C in the Figure in §7.3. Let $\varepsilon(c)$ depend on c but suppose that it takes the same values in each pair in A or in B or in C; the value of $\varepsilon(c)$ will be the contribution to the energy of an arrow configuration. The partition function will be

7.5.1 $$Z(\beta) = \sum_C e^{-\beta E(C)}, \qquad E(C) = \sum_c \varepsilon(c) \tag{7.5.1}$$

where the sum runs over all vertices, *i.e.* over all the arrow configurations compatible with the six-vertex constraint and which can be put on the NM vertices.

Consider first the case $\varepsilon(c) = 0$ (the ice model). Given a configuration C consider the M rows of vertical arrows. We associate a number $\sigma_{mj} = \pm 1$ to the j-th bond ($j = 1, \ldots, N$) on the m-th row ($m = 1, \ldots, M$) indicating whether the arrow points up ($+$) or down ($-$). Let $\underline{\sigma}_m = (\sigma_{m1}, \ldots, \sigma_{mN})$.

Given $\underline{\sigma}_1, \ldots, \underline{\sigma}_M$, *i.e.* the collection of vertical arrow configurations there will be in general several horizontal arrows settings that will be compatible with $\underline{\sigma}_1, \ldots, \underline{\sigma}_M$, *i.e.* which together with the vertical arrows form an allowed configuration around each vertex. Of course the set of horizontal configurations that can be between two rows $\underline{\sigma}_r$ and $\underline{\sigma}_{r+1}$ depends solely upon $\underline{\sigma}_r, \underline{\sigma}_{r+1}$. Hence we can define

7.5.2 $$\begin{aligned} T(\underline{\sigma}, \underline{\sigma}') = \{&\textit{number of horizontal configurations allowed} \\ &\textit{between two rows of vertical arrows } \underline{\sigma} \textit{ and } \underline{\sigma}'\}\,, \end{aligned} \tag{7.5.2}$$

and for instance in this case with $\varepsilon(c) \equiv 0$ it is:

7.5.3 $$Z = \sum_{\underline{\sigma}_1, \ldots, \underline{\sigma}_M} T(\underline{\sigma}_1, \underline{\sigma}_2) \cdot T(\underline{\sigma}_2, \underline{\sigma}_3) \ldots T(\underline{\sigma}_M, \underline{\sigma}_1)\,. \tag{7.5.3}$$

Therefore the free energy (which in the latter example is the so-called *residual entropy*, *i.e.* a number that measures how many configurations are possible for the system) is

7.5.4 $$I = \lim_{N \to \infty} \lim_{M \to \infty} \frac{1}{MN} \log \operatorname{Tr} T^M = \lim_{N \to \infty} \frac{1}{N} \log \lambda_{\max}\,. \tag{7.5.4}$$

One checks, see Fig. 7.3.1, that

7.5.5 $$T(\underline{\sigma}, \underline{\sigma}) = 2, \qquad T(\underline{\sigma}, \underline{\sigma}') = 0, 1, \quad \text{ if } \underline{\sigma} \neq \underline{\sigma}'\,. \tag{7.5.5}$$

If $\underline{\sigma} = (\sigma_1, \ldots, \sigma_N)$ and $\underline{\sigma}' = (\sigma'_1, \ldots, \sigma'_N)$ and if $1 \le x_1 < x_2 < \ldots < x_n \le N$ are the n-labels for which $\sigma_{x_j} = -1$ and $1 \le x'_1 < x'_2 < \ldots < x'_{n'} \le N$ the n'-labels for which $\sigma'_{x'_j} = -1$, then $T(\underline{\sigma}, \underline{\sigma}') = 1$ only when $n = n'$ and one of the following two chains of inequalities holds:

7.5.6
$$1 \le x_1 \le x'_1 \le \ldots \le x_n \le x'_n \le N$$
$$1 \le x'_1 \le x_1 \le \ldots \le x'_n \le x_n \le N\,. \tag{7.5.6}$$

Hence we see that T is divided into blocks which do not mix spaces with different "total spin", *i.e.* one must have $\sum_i \sigma_i = N-2n = N-2n' = \sum_i \sigma'_i$ to have $T(\underline{\sigma}, \underline{\sigma}') \neq 0$.

Fixed n (*i.e.* the total spin) we can consider the matrix $T^{(n)}$ obtained by restricting T to the space generated by the unit vectors that we can label $|\underline{\sigma}\,\rangle$ with $\sum_i \sigma_i = N-2n$. This can therefore be regarded as an operator on the functions $f(x_1, \ldots, x_n)$ with $1 \le x_1 < x_2 \ldots < x_n \le N$:

7.5.7
$$\begin{aligned}(T^{(n)}f)(x_1,\ldots,x_n) = &\sum_{y_1=1}^{x_1}\sum_{y_2=x_1}^{x_2}\ldots\sum_{y_n=x_{n-1}}^{x_n}{}^{*}f(y_1,\ldots,y_n)+\\ &+\sum_{y_1=x_1}^{x_2}\sum_{y_2=x_2}^{x_3}\ldots\sum_{y_n=x_n}^{N}{}^{*}f(y_1,\ldots,y_n)\end{aligned} \tag{7.5.7}$$

where the $*$ denotes that the terms in which there are at least two equal y's must be considered absent, see the following footnote 5.

If $\lambda^{(n)}_{\max}$ is the largest eigenvalue of $T^{(n)}$ then $\lambda_{\max} = \max_{1\le n\le N}\lambda^{(n)}_{\max}$. One can, however, show that in the most interesting cases $\lambda_{\max} = \lambda^{(\frac{N}{2})}_{\max}$.

Bethe's ansatz is that the largest eigenvector of the matrix $T^{(n)}$ is a linear combination of plane waves:

7.5.8
$$f(x_1,\ldots,x_n) = \sum_P A_P\, e^{i\sum_j x_j K_{P_j}} \tag{7.5.8}$$

where the sum runs over the $n!$ permutations $P = (p_1, \ldots, p_n)$ of the n indices and $K_1, \ldots, K_n$ are n *distinct "wave numbers"*.

The idea of trying to find eigenvectors of the form (7.5.8) appeared first in the work of Bethe, [Be31], who found that the eigenvectors of the matrix H defining the one-dimensional *Heisenberg model* could be expressed in that form.[5]

5 The Heisenberg model on the lattice $1, 2, \ldots, N$ with periodic boundary conditions is an operator written in terms of the matrices defined in (7.4.2) and it is:

$$H = \sum_{j=1}^{N} J_x\sigma^x_j\sigma^x_{j+1} + J_y\sigma^y_j\sigma^y_{j+1} + J_z\sigma^z_j\sigma^z_{j+1}, \qquad \sigma_1 \equiv \sigma_{N+1} \tag{$*$}$$

which can be regarded as a matrix acting on the vectors in (7.4.3) which can be denoted $|\underline{\sigma}\,\rangle \stackrel{def}{=} |x_1,\ldots,x_n\,\rangle$ if $x_1,\ldots,x_n$ are the lattice pints where $\sigma_{x_j} = -1$. Thus if the generic vector is written $\sum_{x_1,\ldots,x_n} f(x_1,\ldots,x_n)\,|x_1,\ldots,x_n\,\rangle$ the operator H becomes a matrix acting on the same space as the transfer matrix (7.5.7). If $J = J_x = J_y$ a key remark is that given T the matrix H commutes with T if the coefficient $J_z/J = \Delta$ is suitably chosen as a function of the parameters a, b, c of the six-vertex model: $\Delta = \frac{a^2+b^2-c^2}{2ab}$. Then, see (7.3.4) and (7.3.5), $\Delta = \frac{1}{2}$ for the ice model, $\Delta = 1 - \frac{1}{2}e^{2\beta\varepsilon}$ in the case of the F-model and $\Delta = \frac{1}{2}e^{\beta\varepsilon}$ for the KDP model.

Bethe's eigenvectors were not immediately useful, not even to compute the actual value of the lowest eigenvalue of H (*ground state energy*) because the coefficients A_P were difficult to treat and their evaluation required solving a linear integral equation which, at the time, could not be studied.

Wave functions of the form (7.5.8) turned out to be very useful also in other problems: the "new era" started when it was shown, in [LL63], that the ground state energy of a simple one-dimensional Bose gas with a *nontrivial* interaction corresponded to an eigenvector of the form (7.5.8).

In the papers [YY66] a complete study of the integral equation necessary for the evaluation of the ground state of the Heisenberg model was presented, and in [Li67a] it was discovered that one could find eigenvectors of (7.5.7) of the form (7.5.8), among which one with $f(x_1,\ldots,x_n) > 0$ corresponding to $n = N/2$ when N is even, as usually assumed. From this Lieb was able to show, in the ice model case, that the largest eigenvalue of T was precisely the same as that which gave the ground state energy of a corresponding Heisenberg model (with suitably chosen couplings, see footnote [5]) and whose form had become known by the [YY66] key work.

The commutation property, cited in footnote [5], between the ice model transfer matrix and the Heisenberg model Hamiltonian with $\Delta = \frac{1}{2}$ was also a consequence of the results on the ice model. This becomes clear as soon as one realizes that the two matrices have the same eigenvectors (all of the Bethe ansatz form, (7.5.8)).

The knowledge of the ground state eigenvalue for the ice model transfer matrix led to the explicit evaluation of the ice model residual entropy and, shortly afterwards, the works [Li67b],[Li67c], and [Su67] determined the largest eigenvalue of the matrix T in the F and KDP-models.

It would be fairly easy to reproduce the work of [Be31] to see that the matrix H, written in the basis of the last footnote [5], admits eigenvectors of the form (7.5.8). And remark that the Heisenberg model matrix H and transfer matrix T for the above considered $6V$ models commute (if, of course, Δ is suitably chosen) implies that T has eigenvectors of the form (7.5.8) and provides a way for finding the solution oof the model: *this, however, was not the path followed* in the discovery of the ice model solution.

A direct check that one can adjust A_P, K_P in (7.5.8) to make them eigenvectors of the matrix (7.5.7), [Li67a], is possible and very instructive, although it is surprisingly difficult to write in words. The procedure suggested in [Ba82] is perhaps the simplest.

Having noted that T decomposes into blocks that do not mix vectors in the spaces generated by the basis elements $|x_1,\ldots,x_n\rangle$ (*i.e.* the functions of $x'_1 < \ldots < x'_m$ vanishing unless $m = n$ and $x'_i = x_i$, when their value is 1, see footnote [5]) with different n's one studies first the case $n = 0$ (trivial), then $n = 1$ (also very easy), then $n = 2$ which is easy but requires attention as the algebra is already quite involved. The calculation is strongly recommended, and one should first attempt it in the ice model case $\Delta = \frac{1}{2}$.

For the ice model (of the F and KDP-models which are not harder) one

finds

7.5.9
$$\frac{A_{12}}{A_{21}} = -\frac{1-2\Delta z_1 + z_1 z_2}{1-2\Delta z_2 + z_1 z_2}, \qquad z_j = e^{iK_j}$$
$$z_1^N = \frac{A_{12}}{A_{21}}, \qquad z_2^N = \frac{A_{21}}{A_{12}}\,. \tag{7.5.9}$$

If one has *really* performed the calculation at least in the $\Delta = \frac{1}{2}$ case, with the necessary patience, then: "*the solution of the eigenvalue problem for arbitrary n is a straightforward generalization of the $n = 2$ case*", p.138 of [Ba82]! If $P = (p_1, \ldots, p_n)$ is the generic permutation of $1, \ldots, n$, then:

$$A_{p_1,\ldots,p_n} = (-1)^P \prod_{i<j} s_{p_i,p+j}, \qquad s_{p,q} \stackrel{def}{=} 1 - 2\Delta z_q + z_p z_q,$$

$$z_j^N = (-1)^{n-1} \prod_{\ell\neq j} \frac{s_{\ell,j}}{s_{j,\ell}} (7.5.10)$$

where the last relations must be interpreted as equations for z_j hence, by (7.5.9), for the K_j's. The calculations also provide the value of the eigenvalue λ corresponding to a sequence $K_1, \ldots, K_n$ of n values that are pairwise distinct and satisfy the last of (7.5.10). For instance in the ice model case:

7.5.11
$$\lambda = \prod_{j=1}^n \frac{1}{1-z_j} + \prod_{j=1}^n \frac{z_j}{1-z_j} \tag{7.5.11}$$

It is remarkable that if $|\Delta| < 1$ and p, q are real then

7.5.12
$$\frac{s_{p,q}}{s_{q,p}} = -\frac{1 - 2\Delta e^{ip} + e^{i(p+q)}}{1 - 2\Delta e^{iq} + e^{i(p+q)}} = e^{i\Theta(p,q)}\,. \tag{7.5.12}$$

where the Θ's are real and are given by

7.5.13
$$\Theta(p,q) = 2\,\mathrm{arctg}\,\frac{\Delta \sin\frac{1}{2}(p-q)}{\cos\frac{1}{2}(p+q) - \Delta\cos\frac{1}{2}(p-q)} \tag{7.5.13}$$

which is real if p, q are real. The conditions on K_j become

7.5.14
$$2\pi I_j = N K_j + \sum_{\ell=1}^n \Theta(K_j, K_\ell), \qquad I_j = j - \frac{1}{2}(n+1) \tag{7.5.14}$$

which has been discussed in [YY66] and shown to admit, for all $|\Delta| < 1$, a unique real solution $K_1, \ldots, K_n$. Therefore, modulo mathematical rigor problems, we expect that $K_1, \ldots, K_n$ *become dense* when $N \to \infty$ and $n/N = \delta$ stays constant; the number of K_j's in an interval dK should be described by a density $\rho_\delta(K)$ such that $\int_{-\pi}^{\pi} \rho_\delta(K) dK = \delta = \frac{n}{N}$. The distribution $\rho_\delta(K)$ will be nonzero inside an interval $[-Q, Q] \subseteq [-\pi\delta, \pi\delta]$ (because j in (7.5.14) varies between 1 and $\frac{1}{2}(n+1)$).

The form of (7.5.14) means that the number of K_j's which are $\le K$ is such that

7.5.15
$$N\int_{-Q}^{K}\rho_\delta(K')dK' = \pi(n+1)+NK+N\int_{-Q}^{Q}\Theta(K,K')\rho_\delta(K')dK'; \quad (7.5.15)$$

hence differentiating with respect to K:

7.5.16
$$2\pi\rho_\delta(K) = 1+\int_{-Q}^{Q}\partial_K\Theta(K,K')\rho_\delta(K')\,dK' \qquad (7.5.16)$$

which is an equation whose solution for $\delta\neq\frac{1}{2}$ is still an open problem.

The problem for $\delta=\frac{1}{2}$ was solved in 1938 (reference to Hulthén, [Hu38], in [Ba82]) for the case $\Delta=-1$, and in [Wa59] for the cases $\Delta<-1$. The $|\Delta|<1$ cases were completely solved by Yang and Yang, [YY66], for $\Delta<1$.

The case $\Delta>1$ is trivial, as noted by Lieb, because one can see that the maximum eigenvalue is the one corresponding to $n=0$ and it is a^N+b^N, see [Ba82], (see (7.3.1) for the definition of a,b).

The reason why the case $n=\frac{1}{2}N$ is so special and exactly computable lies in the existence of a change of variables transforming (7.5.16) into a convolution equation, in the cases $-1<-\cos\mu\overset{def}{=}\Delta<1$. The change of variables is $K\longleftrightarrow\alpha$:

7.5.17
$$e^{iK}=\frac{e^{i\mu}-e^{\alpha}}{e^{i\mu+\alpha}-1} \qquad (7.5.17)$$

or

7.5.18
$$\frac{dK}{d\alpha}=\frac{\sin\mu}{\cosh\alpha-\cos\mu} \qquad (7.5.18)$$

which maps the interval $[-Q,Q]$ into $(-\infty,\infty)$ and $2\pi\rho_{\frac{1}{2}}(K)\longleftrightarrow R(\alpha)$ with $R(\alpha)$ satisfying:

7.5.19
$$R(\alpha)=\frac{\sin\mu}{\cosh\alpha-\cos\mu}-\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\sin 2\mu}{\cosh(\alpha-\beta)-\cos 2\mu}R(\beta)d\beta \qquad (7.5.19)$$

which can be solved by Fourier transform, and it even leads to the simple solution

7.5.20
$$\hat{R}(x)=\frac{1}{2}\frac{1}{\cosh\mu x} \qquad (7.5.20)$$

for the Fourier transform $\hat{R}$ of R!!

The first instance in which a transformation of the type (7.5.19) is used to transform the integral equation (7.5.16) into a simple equation is in the remarkable brief paper [Wa59] which introduced the change of variables corresponding to (7.5.17) in the case $\Delta<-1$. The latter paper is based upon another remarkable paper, [Or58], which studies the same equation in an approximate way.

It is now a matter of simple algebra to obtain the formulae of §7.4, (7.3.3),(7.3.4),(7.3.5), see [LW72],[Ba82].

The $8V$ model is *not* soluble by Bethe's ansatz; I cannot give here even a sketchy account of its solution. The reader should look at Baxter's book, [Ba82], detailing what is one the main achievements of mathematical physics in the 1970s. The book also illustrates several other exactly soluble models.

Chapter *VIII*

Brownian Motion

.

§8.1. Brownian Motion and Einstein's Theory.

Brownian motion was first observed by Brown (1828): he recognized that the motion of "molecules", of size of $\sim 10^{-3}\, cm$, of a pollen in a solution ("colloid") was not due to internal causes, as it was believed at the time, but it had a mechanical (unknown) origin. It was in fact observable by looking at particles of comparable size and of *any* material, both organic and inorganic.

Very soon the movements were attributed to collisions with the microscopic constituents of matter. Among the first to recognize this was Cantoni, [Ca67], see p.93 in [Pa82], in a remarkable paper he says

"in fact, I think that the dancing movement of the extremely minute solid particles in a liquid, can be attributed to the different velocities that must be proper at a given temperature of both such solid particles and of the molecules of the liquid that hit them from every side. I do not know whether others did already attempt this way of explaining Brownian motions..."

In this paper an impressive number of experiments, performed by Cantoni himself, are reported in which he finds evidence for the equipartition of energy between the suspended particles and the solvent molecules to conclude:

"In this way Brownian motion provides us with one of the most beautiful and direct experimental demonstrations of the fundamental principles of the mechanical theory of heat, making manifest the assiduous vibrational state that must exist both in liquids and solids even when one does not alter their temperature".

This work is most remarkable also in view of the fact that it is *contemporary* of the first papers of Boltzmann on the heat theorem and equipartition.

Brownian motion attracted the interest of many leading scientists, among which was Poincaré. Brownian motion theory was worked out by Einstein and, independently, by Smoluchowski, (1905-1906), soon followed by the experimental confirmation of Perrin, (1908), see [Ei56],[VS06];[Pe70].

The main critique (Nägeli, [VN79]) to the microscopic kinetic nature of Brownian motion was the remark that experimental data and kinetic theories permitted one to estimate that (the particles in suspension being hundreds of millions times larger than the molecules of the liquid) the velocity variations at each collision had a random sign; so that it seemed inconceivable that one could see a nonvanishing average effect. A fallacious argument, as it was stressed (for instance) by Poincaré (1904), [Po00]. He also noted, with others, that the hypothesis that the colloidal particles motion had a kinetic nature could contradict thermodynamics (see below).

The fallacy of the reasoning was, in any event, well known as it appears

from an esoteric article by Bachelier (1900), [Ba00], on the austere *Annales de l'Ecole Normale* of Paris (a few pages away from the French translation of Hilbert's *Grundlagen der Geometrie*); it is a paper on stock market speculations where, precisely, the problem is posed of how can many random small variations produce visible effects (see below).

Einstein's approach starts with the remark that the suspended particles, although of huge size, can still be considered as large molecules and, hence, one can apply statisical mechanics to them, so that they will exercise osmotic pressure, just as with ordinary solutions, satisfying (therefore) Raoult-van t'Hoff's law at least at small concentrations. Hence Van t'Hoff's law holds not only for solutions of microscopic particles but also for calculating the partial pressure due to particles of arbitrary sizes (*e.g.* glass balls).

The idea was "revolutionary" and, as Einstein realized, possibly in contrast with classical thermodynamics, but *not* with statisical mechanics and the atomic hypothesis. Hence he immediately posed the question of how to find observable macroscopic consequences.

Nonrectilinear motion of particles is thus attributed to their random collisions with molecules. Hence it is a random motion, at least when observed on time scales τ large compared to the time necessary to dissipate the velocity v acquired in a single collision with a molecule (by friction, due, also to microscopic collisions between fluid molecules). The dissipation of such velocity can be estimated, in the case of macroscopic particles, by remarking that in a single collision with a molecule the acquired speed v is dissipated into heat by the action of a force F which, by *Stokes' law*, is

8.1.1
$$m\frac{\partial v}{\partial t} \equiv F = -6\pi\eta R v \tag{8.1.1}$$

where η is the fluid *viscosity coefficient*, R is the radius of the suspended particles, and v the speed; hence the characteristic time scale for the loss of the velocity acquired in a single collision is $t_0 = (6\pi\eta m^{-1}R)^{-1}$. This is a very short time (for instance if $R = 1\,\mu$, and if m is evaluated by assuming that the density of the material constituting the large particles is the same as that of the liquid in which they are suspended (*i.e.* water, so that $\eta = 10^{-3}$ cgs-units), one realizes that the time scale is $t_0 \approx 10^{-7}$ sec). Therefore on the time scale $\tau \gg t_0$ motion will be *diffusive*. In such motions there is transport of matter only when there is a density gradient.

The logic of Einstein's analysis is quite fascinating. Using an ideal experiment (a method characteristic of his thinking) he links microscopic quantities to macroscopic ones. The background solvent fixes the temperature and the time scale over which a particle undergoes a diffusive motion: to find the diffusion coefficient for a *single* particle one considers a gas of particles of *arbitrary* density ν, but so small that the ideal osmotic pressure law holds. We recall *Raoult–van t'Hoff's law*: if p is the osmotic pressure (*i.e.* the partial pressure due to the particles) and ν is their numerical density one has $p = k_B T\nu$, with T being the fluid temperature and k_B Boltzmann's constant. This is done in spite of the fact that in the classical experiments the

suspended particles are so few that they can be considered isolated from each other.

Hence in the first step of the ideal analysis one replaces a single colloidal particle with a gas of such particles, with density ν.

One then imagines that some external force F acts on the gas (also a fiction, in the ideal experiment) which only acts upon the colloidal particles (and not on the solvent fluid); it generates, in a stationary state, a gradient $F\nu$ of pressure, denoted $\partial_x p$. The pressure gradient also equals $k_B T \partial_x \nu$ by the Raoult–van t'Hoff law (or more precisely by the assumption that the osmotic pressure law holds for macroscopic particles). The pressure gradient which exactly balances the force action is $\partial_x p = F\nu = k_B T \partial_x \nu$ by the Raoult–van t'Hoff law.

Supposing that the solvent obeys the Navier–Stokes equations one can then compute, via Stokes' law, the particle velocity in terms of the viscosity (strictly speaking here it is necessary that particles be in fact macroscopic) so that one can compute the flux generated by the pressure gradient:

8.1.2 $$\Phi = \nu\, v = -\frac{\nu\, F}{6\pi\eta R} = -\frac{k_B T}{6\pi\eta R}\,\partial_x \nu\,. \tag{8.1.2}$$

Finally the assumption that the individual particles undergo a diffusive motion implies that the flux has to be proportional to the density gradient giving:

8.1.3 $$\Phi = -D\,\partial_x \nu \tag{8.1.3}$$

where the proportionality constant D is the *diffusion coefficient*.

Equating the two expressions for the flux of particles all auxiliary quantities, used to mount the ideal experiment, have disappeared and one infers that assuming kinetic theory then a macroscopic particle (even just one) in a fluid and in thermal equilibrium (*i.e.* in a stationary state) must have a diffusive motion with a diffusion constant related to the viscosity by

8.1.4 $$D = \frac{k_B T}{6\pi\eta R}, \tag{8.1.4}$$

which is called the *Einstein–Smoluchowski relation*: that one should attribute entirely to Einstein, see the following §8.2.

The quantity D is also directly related to the average value (over many trajectories) $\langle \underline{r}(t)^2 \rangle$ of the squared displacement $\underline{r}(t)^2$ of the colloidal particle in a time interval t; we shall see that $\langle \underline{r}(t)^2 \rangle = 6Dt$. Since the value of $\langle \underline{r}(t)^2 \rangle$ is directly measurable in a microscope, this is a first theoretical relation that can be checked experimentally.

Conceiving of macroscopic particles as behaving like microscopic molecules (and generating an osmotic pressure obeying Raoult–van t'Hoff law, much as true chemical solutions do) is an important idea that was, in itself, a novelty brought by Einstein's work (heralded by Cantoni's experiments, [Ca67]). It allowed everybody who had not yet accepted the atomic hypothesis to see

that thermodynamics laws *must* have a statistical nature. This in fact happened at least after Perrin showed, experimentally, that Brownian motion theory was correct, [Pe70].

In fact, since it seems possible to build semi-permeable walls for macroscopic objects, it becomes possible to perform thermodynamic cycles at constant temperature which, by using osmotic pressure, convert heat into work: the walls create an entity similar to Maxwell's daemon. One can think of a cylinder filled with liquid and divided in two by a semi-permeable movable wall; on the left side there is also a colloidal solution seeing the impermeable side of the wall, while the right side contains no colloid. The wall can then be pushed to the right by using the osmotic pressure to extract work from the process; at the end of the run the wall is taken out (performing a work as small as we wish, in principle, because we only need to displace the wall horizontally) and reinserted back at the original position but with the two faces inverted.

Then comes a long waiting time while the observer does nothing but witness the colloidal particles randomly hit the wall on the permeable side and get caught in the left part of the cylinder, again. After the last colloidal particle crosses the wall the initial conditions are restored and Carnot's principle has been violated.

The infinitely sharp eye of Maxwell's daemon can thus be replaced by our microscope, as Poincaré stressed (having in mind a somewhat different apparatus based on the same ideas). Perrin highlighted this same aspect of the Brownian motion phenomenon, and he also noted that a machine like the above would have required unimaginably long times to extract appreciable amounts of energy, see §51 of [Pe70].

It is important to keep in mind that here we are somewhat stretching the validity of thermodynamic laws: the above machines are very idealized objects, like the daemon. They cannot be realized in any practical way: one can arrange them to perform one cycle, perhaps; (and even that will take forever if we want to get an appreciable amount of energy, see §51 of [Pe70]), but what one needs to violate the second law is the possibility of performing as many energy producing cycles as required (taking heat out of a single reservoir). *Otherwise their existence "only" proves* that the second law has only a statistical validity, a fact that had been well established since the work of Boltzmann.

In fact an accurate analysis of the actual possibility of building walls semi-permeable to colloids and of exhibiting macroscopic violations of the second principle runs into grave difficulties: it is not possible to realize a *perpetual motion* of the second kind by using the properties of Brownian motion. It is in fact possible to obtain a single violation of Carnot's law (or a few of them), of the type described by Perrin, but as time elapses and the machine is left running, isolated and subject to physical laws with no daemon or other ideal extraterrestrial being intervening (or performing work unaccounted for), the violations (*i.e.* the energy produced per cycle) vanish because the cycle will be necessarily performed as many times in one direction (apparently

violating Carnot's principle and producing work) as in the opposite direction (using it).

This is explained in an analysis of Feynman, see [Fe63], vol. 1, §46, where the semi-permeable wall is replaced by a wheel with an anchor mechanism, a "*ratchet and a pawl*", allowing it to rotate only in one direction under the impulses communicated by the colloidal particle collisions with the valves of a second wheel rigidly bound to the same axis. Feynman's analysis is really beautiful, and remarkable as an example of how one can still say something nonboring about perpetual motion. It also brings important insights into the related so-called "reversibility paradox" (that microscopic dynamics generates an irreversible macroscopic world).

Diffusive motion produces a displacement $\underline{r}(t)$ over a time t whose squared average is $\langle r^2\rangle(t) = 6Dt$, because the probability $f(\underline{x},t)d^3\underline{x}$ for finding a particle, initially located at the origin, in the little cube $d\underline{x}$ around $\underline{x}$ is the solution of the diffusion equation $\partial_t f(\underline{x},t) = D\Delta f(\underline{x},t)$, *i.e.*

8.1.5 $$f(\underline{x},t) = \frac{e^{-\underline{x}^2/4Dt}}{(4\pi Dt)^{3/2}} \tag{8.1.5}$$

(the equation that Einstein derives by imitating Boltzmann's method to obtain the Boltzmann's equation also finding, at the same time, a microscopic expression for the diffusion coefficient D). The squared average value of the displacement is then simply:

8.1.6 $$\langle \underline{r}(t)^2\rangle = \int \underline{x}^2 f(\underline{x},t)\, d^3\underline{x} = 6D\,t \tag{8.1.6}$$

We see that, although each collision produces a very small velocity variation, immediately followed by variations of similar size and of either sign, nevertheless the particle undergoes a motion that over a long time (compared to the frequency of the collisions) leads to a change of each coordinate of the order of $\sqrt{2Dt}$ (or $\sqrt{6Dt}$ if one looks at the three-dimensional variation) which not only is nonvanishing, but can also be considerably large and observable.

As an application Einstein deduced (1906) the value of Boltzmann's constant k_B, hence of Avogadro's number N_A, from the measured diffusion of sugar suspended in water, finding $N_A = 4.0\times 10^{23}$: the error being mainly due to a computational mistake. On the basis of accurate experiments, by using the theory of Einstein, Perrin and collaborators obtained a value essentially equal to the recently accepted value of N_A, see §77 of [Pe70].

Brownian motion theory was derived by Einstein without him being really familiar with the details of the experiments that had been performed for about 80 years. He proceeded deductively, relying on ideal experiments, starting from the remark that particles, *even if of macroscopic size*, *had to obey* the laws of statisical mechanics. In particular they had to show energy equipartition and their osmotic pressure had to obey the perfect gas law (Raoult–van t'Hoff's law).

His theory leads manifestly to motions that, if observed over time scales long compared to t_0 (cf. lines following (8.1.1))[1] must be motions for which the velocity would depend on the time interval over which it is measured and it would diverge in the limit $t \to 0$ or, better, it would become extremely large and fluctuating as t approaches the time scale t_0 beyond which the theory becomes inapplicable.

It provided, therefore, an example of an actual physical realization of certain objects that had until then been just mathematical curiosities, like continuous *but nondifferentiable* curves, discovered in the '800s by mathematicians in their quest for a rigorous formulation of calculus; Perrin himself stressed this point very appropriately, see §68 of [Pe70].

The assumption that fluid resistance to macroscopic particle motion follows Stoke's law is by no means essential but it is a characteristic aspect distinguishing Einstein's theory from Smoluchowski's, as we shall see below. In fact if the assumption was changed to $v = C_R F$ with C_R suitably depending on R, then (8.1.4) would be replaced by $D = kTC_R$. If, for instance, the particle was suspended in a rarefied gas, rather than in an incompressible liquid, then C_R would be different.

More precisely if the colloidal particle proceeds with velocity v in a gas with density ρ, then the number of gas particles colliding with an average velocity $-v_m$ is $\pi R^2(v+v_m)\rho/2$, while $\pi R^2(v-v_m)\rho/2$ is the number of particles colliding with an average velocity $+v_m$. The former undergo a momentum variation, per unit time, $2m(v_m+v)$ and the latter $2m(v_m-v)$. Hence, instead of Stokes' law, the force of the fluid on the particle is

8.1.7
$$\frac{\rho}{2}\pi R^2[(v+v_m)^2-(v_m-v)^2)2m = c' R^2 v_m m\rho v \tag{8.1.7}$$

with $c' = 4\pi$.

In the above calculation we supposed that half of the particles had velocity equal the absolute velocity average and half an opposite velocity; furthermore the particle has been treated as a disk perpendicular to the direction of motion. A more correct treatment should assume a Maxwellian velocity distribution and a spherical shape for the particle. The evaluation of the corrections is without special difficulties if one assumes that the gas is sufficiently rarefied so that one can neglect the *recollision phenomena* (*i.e.* repeated collisions between the particle and the *same* gas molecule, and it leads to a final result identical to (8.1.7) but with a different factor replacing 4π. One would eventually find, following the argument leading to (8.1.4),

8.1.8
$$D = \frac{k_B T}{c' R^2 m v_m \rho} = \frac{k_B T}{cR^2\rho\sqrt{2mk_B T}} = \frac{\sqrt{k_B T}}{cR^2\rho\sqrt{2m}} \tag{8.1.8}$$

and the constant c is in fact $2\sqrt{\pi}$. For obvious reasons the regime in which the expression (8.1.7) for the friction holds is called *Doppler's regime* and

[1] *i.e.* long compared to $1\,\mu sec$ as is necessarily the case because of our human size

it is relevant for rarefied gases, while the *Stokes regime*, in which (8.1.1) is acceptable, pertains to friction in liquids.

§8.2. Smoluchowski's Theory

Smoluchowski's theory, shortly following that of Einstein, shines light on Einstein's hypotheses. The latter lead to (8.1.5) at the price of tacit assumptions similar to the *molecular chaos* hypothesis familiar from Boltzmann's attempts at deriving Boltzmann's equation: in fact according to (8.1.5) colloidal particles show a diffusive motion, with mean square displacement proportional to time t.

Smoluchowski, to disprove Nägeli's argument, considers a concrete microscopic model for the collisions: a particle of mass M is subject to a large number of collisions with the molecules, of mass m, of the fluid ($\approx 10^{16}\,\mathrm{sec}^{-1}$ in many cases). If $\underline{v}_k$ is the particle velocity after k collisions and if the k-th collision is with a molecule with velocity $\underline{v}$ before colliding, one infers from the elastic collision laws that:

8.2.1
$$\underline{v}_{k+1} \approx \underline{v}_k + \frac{m}{M}(\mathcal{R}-1)\,\underline{v} \qquad \text{if } |\underline{v}| \gg |\underline{v}_k|,\ M \gg m \tag{8.2.1}$$

where $\mathcal{R}$ is a random rotation (depending on the impact parameter, also random). The above equation only deals with a single collision between two particles and it does not take into account that the heavy particle moves in a gas of light particles with positive density: this causes a cumulative friction effect; hence when the velocity $\underline{v}_k$ grows the Doppler friction in (8.1.7) will start to damp it by a force $-c'R^2 v_m \rho m \underline{v}_k$, hence with acceleration $-\lambda \underline{v}_k$ with $\lambda = c'R^2 v_m \rho m/M$. Thus the velocity variation should rather be as

8.2.2
$$\underline{v}_{k+1} \approx \underline{v}_k e^{-\lambda\tau} + \frac{m}{M}(\mathcal{R}-1)\,\underline{v} \tag{8.2.2}$$

which includes, empirically, the damping effect.

We can consider $(\mathcal{R}-1)\underline{v}$ as a random vector, at each collision, with zero average and square width $(m/M)^2\langle((\mathcal{R}-1)\underline{v})^2\rangle = 2(m/M)^2 v_m^2$. Furthermore collisions take place, in average, every time interval τ such that

8.2.3
$$\tau\pi R^2 \rho v_m = 1, \quad \rightarrow \quad \lambda\tau = c'\frac{m}{M} \tag{8.2.3}$$

hence the space run in a time t during which $n = t/\tau$ collisions take place is (if the initial positon and velocity of the particle are both $\underline{0}$, for simplicity)

8.2.4
$$\underline{r} = \sum_{k=0}^{n-1} \tau \sum_{h=0}^{k} e^{-\lambda\tau(k-h)} \underline{w}_k \tag{8.2.4}$$

with $\underline{w}_k$ independent random vectors with $\langle \underline{w}_k^2 \rangle = 2v_m^2(m/M)^2$. Hence we

can immediately compute the average square dislacement as[2]

8.2.5
$$\langle\frac{r^2}{t}\rangle = 2v_m^2\Big(\frac{m}{M}\Big)\tau^2\frac{1}{(\lambda\tau)^2}\frac{1}{\tau} = \frac{2v_m^2}{c'^2}\frac{1}{\pi\rho R^2 v_m} = 6a\frac{k_BT}{c'R^2m\rho v_m} \qquad (8.2.5)$$

where $a = 2/(c'\pi)$ is a numerical constant of $O(1)$ (see below) and c' was introduced in (8.1.7); we see that we find again, essentially, Einstein's formula in the case in which Stokes law is replaced by Doppler's resistance of a rarefied gas, just as we would expect to find given the nature of the model (apart from the factor a). One shall realize, see §8.3, that Uhlenbeck–Ornstein's theory is a *macroscopic version* of Smoluchowski's theory.

This led Smoluchowski to say that if, instead, the fluid is an incompressible liquid one has "simply" to replace, in the denominator of (8.2.5), the Doppler regime viscosity with that in the Stokes regime, with a somewhat audacious logical leap; he thus finds:

8.2.6
$$D_{smol.} = aD_{einst.} \qquad (8.2.6)$$

Hence Smoluchowski's theory is in a sense more ambitious than the Einsteinian theory because it attempts at *proving* that the colloid motion *is* a diffusive one without neglecting completely the time correlations between consecutive collisions (which Einstein, as already mentioned, implicitly neglects). The model proposed is to think of the fluid as a rarefied gas which, therefore, does not obey the Stokes viscosity law. Strictly speaking Smoluchowski's model deals with a colloid realized in a rarefied gas, a situation not very relevant for the experiments at the time, because it is not applicable to a colloid realized in a fluid. Einstein's method is more general and applies to both cases, although it does not really provide a microscopic justification of the diffusive nature of the motions.

Conceptually Smoluchowski could not possibly obtain Einstein's formula because he was not able to produce a reasonable microscopic model of a fluid in the *Stokes regime* (which even today does not have a satisfactory theory). His method in fact is not very "objective" even in the rarefied gas case since it leads to a result for D affected by an error of a factor a with respect to Einstein's.

This factor can be attributed to the roughness of the approximations, mainly to the not very transparent distinction between velocity and average velocity in the course of the derivation of (8.2.5), which does not allow us to compute an *unambiguously correct* value for a. Nevertheless Smoluchowski, without the support of the macroscopic viewpoint on which Einstein was basing his theory, is forced to take seriously the factor a that he finds and to transfer it (with the logical jump noted above) to an incorrect result in the case of a liquid motion.

---

[2] We use that $\frac{3}{2}\frac{k_BT}{m} = \frac{1}{2}v_m^2$. The factor a changes if one makes a less rough theory of the Doppler friction taking into account that there are differences between various quantities identified in the discussion, like $\sqrt{\langle(\Delta\underline{v}_k)^2\rangle}$ and $\langle|\Delta\underline{v}_k|\rangle$ (which, by the Maxwellian distribution of the velocities, modifies c' hence a).

As in the case of the factor c in (8.1.8) a more precise theory of the collisions between molecules and colloid is possible, in which one replaces the average values of the velocity with fluctuating values (distributed according to the appropriate Maxwellian): in this way the value $a = 1$ arises. Had Smoluchowski proceeded in this manner, although finding the correct result in Doppler regime, *he would have still needed a logical jump* to treat the case of a colloid in a liquid solution.

It thus appears that Smoluchowski's theory *was not comparable* with the experimental data available at the time; this was so for *intrinsic reasons* and it, perhaps, explains why he did not publish his results before Einstein's papers (results that, as he says, he had obtained years earlier). It is not impossible that by reading Einstein's memoir he could make the above analyzed logical jump which was necessary in order to make a comparison of the theory with the experiments (which occupies few lines in his long memoir).

Later Smoluchowski abandoned the factor a and he adopted "Einstein's' value" ($a = 1$).

It remains true, however, that Smoluchowski's work is a milestone in kinetic theory and his was among the first of a series of attempts aimed at obtaining equations for macroscopic continua. The continua are regarded as describing microscopic motions observed over time scales (and space scales) very large compared to microscopic times (and distances), so that the number of microscopic events involved in an observation made on a macroscopic scale was so large that it could be treated by using probability theory techniques (or equivalent methods).

The use of probability theory is the innovative feature of such theories: already Lagrange, in his theory of the vibrating string, imagined the string as composed of many small coupled oscillators: but his theory was entirely "deterministic", so much as to appear artificial.

In 1900, six years before Smoluchowski's work, Bachelier published the above mentioned research, [Ba00], with the rather unappealing title of *Théorie de la spéculation* which, as is maintained by some historians, would have been left unappreciated because it was superseded or shadowed by Einstein's 1905 paper. Bachelier's work, it is claimed, did in fact present the first theory of Brownian motion.

It is in fact only *a posteriori* possible to see a connection between the theory of fluctuations of erratic ("?") stock market indicators and Brownian motion; nevertheless Bachelier's memoir can perhaps be considered to be the first paper in which dissipative macroscopic equations are rigorously derived from underlying microscopic models.

In his work Brownian motion is not mentioned and his model for the evolution of list prices is that of a random increase or decrease by an amount Δx in a time Δt with equal probability. The novelty with respect to the classical error analysis is that one considers the limit in which Δx and Δt tend to 0 while studying the list price variations at various *different times* t under the assumption that they are given by partial sums of the price

variations. In classical error theory one studies only the *total* error, *i.e.* the sum of all the variations, so that the summation index does not have the interpretation of time value, but is an index enumerating the various error values.

One deduces that the probability distribution of the list price values at time t satisfies a diffusion equation; furthermore the probability distribution of successive increments of value is a product of independent Gaussian distributions and one arrives at a kind of preliminary version of the stochastic process that was later studied by Wiener (essentially in [Ba00] "only" the analysis of the continuity of the sample paths as functions of time is missing).

Thus we can consider Bachelier's work as similar to Smoluchowski's theory and, therefore, rather loosely related to Einstein's theory and, furthermore, in his analysis no mention appears of physics and thermodynamics. But we have seen in the above discussion that it is precisely here that one of the main difficulties of Brownian motion lies.

Smoluchovsky's point of view is by no means superseded: in the last few decades it has been developed into very refined theories aiming at understanding the more general problem of deriving macroscopic continua equations from microscopic dynamics: see [Sp91], part II, for a perspective and technical details.

§8.3. The Uhlenbeck–Ornstein Theory

As remarked by Einstein (as well as by Smoluchowski) Brownian motion theory held for experimental observations taking place at time intervals spaced by a quantity large compared to the time scale characteristic for the loss of the velocity acquired in a single collision, which is $t_0 = (6\pi\eta m^{-1}R)^{-1}$.

For shorter time intervals it still makes sense to define the velocity of the particles and motion cannot be described by the diffusive process characteristic of Brownian fluctuations proper. The trajectories appear, when observed over time scales larger than t_0, erratic and irregular so that if one tries to measure the velocity by dividing the space run by the corresponding time one finds a result depending on the time interval size and that becomes larger and larger the more the time interval is reduced. This is an immediate consequence of the fact that, on such time scales, the average of the absolute value of the displacement is proportional to $\sqrt{t}$, rather than to t. But this "divergence" of the velocity *ceases* as soon as one examines motion on time scales short compared to t_0

One is then faced with the problem of developing a theory by describing motions in the "normal" phase at small time intervals, *as well as* in the Brownian phase, at larger time intervals. Langevin proposed a very simple mathematical model for the complete Brownian motion equations.

He imagined that successive collisions with fluid molecules had an effect on the variations of each velocity component that could be described in terms of a random impulsive force $F(t)$ and, hence, the equation of motion of a

coordinate of a colloidal particle would be:

8.3.1
$$m\dot{v} = -\lambda v + F(t) \tag{8.3.1}$$

where λ is the friction coefficient for the colloid motion (*i.e.* $6\pi\eta R$ in the case of a fluid in a Stokes regime and $cR^2 v_m m\rho$ in the case of a Doppler regime, see (8.1.7)).

The Langevin equation, (8.3.1), can be discussed once a suitable random force law is assigned for F. The model proposed by Uhlenbeck and Ornstein for $F(t)$ was that of a *white noise*, *i.e.* that F was such that:

(1) no correlation existed between the values of $F(t)$ at different time instants,

(2) the distribution of an n-tuple $F(t_1), F(t_2), \ldots, F(t_n)$ of values of the force, observed at n arbitrary instants $t_1 < t_2 < \ldots < t_n$, was a Gaussian distribution, and

(3) the average value of $F(t)$ vanished identically as a function of t.

This leads to the notion of a (centered) Gaussian stochastic process and to the more general notion of *stochastic process* and, also, it leads to the possibility of regarding Brownian motion as an "*exactly soluble*" stochastic process.

Consider a stochastic process, *i.e.* a probability distribution, on a space of events that can be represented as functions of one (or more) zero average variables $t \to F(t)$, see §5.7, footnote 9. It is characterized by giving the probability of observing an n–tuple $F(t_1), F(t_2), \ldots, F(t_n)$ of force values when measuring the $F(t)$ at n instants $t_1 < t_2 < \ldots < t_n$ as a Gaussian distribution on the force values.

It can be shown that such a process (*i.e.* the probability distribution of the functions $t \to F(t)$) is uniquely determined by the *two-point correlation function*, also called *the covariance*, or *propagator*. This function is defined as the average value of the product of the function values at two arbitrary instants t_1, t_2:

8.3.2
$$C(t_1, t_2) = \langle F(t_1) F(t_2) \rangle \tag{8.3.2}$$

and this means that the Gaussian distribution of the probability of an arbitrary n–tuple of values of F at n distinct time instants can be simply

expressed in terms of the covariance C (it is, in fact, simply expressible in terms of the inverse matrix of the matrix $C(t_i,t_j)$, $i,j=1,2,\ldots n$).[3]

In this way the "white noise" is defined as the Gaussian process with covariance

8.3.3
$$C(t,t')=f^2\delta(t-t') \tag{8.3.3}$$

where f^2 is a constant and δ is Dirac's delta function.

We shall write f^2 as $2\lambda k_BT$ (thus defining T) and it will appear that if T is identified with the absolute temperature of the solvent then the Uhlenbeck–Ornstein theory and the Einstein theory will agree where they should, see below.

For C given by (8.3.3) an explicit solution of (8.3.1) is possible and it follows, as shown by Uhlenbeck and Ornstein, [UO30], that each velocity and position component which is generated from initial data s_0,v_0 (respectively for the position and the velocity) is a Gaussian process with nonzero average. If $\beta=\lambda/m\equiv t_0^{-1}$ (see the lines following (8.1.1)), their average, at time t, is given by

8.3.4
$$\overline{s}(t)=s_0+\frac{v_0}{\beta}(1-e^{-\beta t}),\qquad \overline{v}(t)=v_0e^{-\beta t} \tag{8.3.4}$$

which follow simply by averaging (8.3.1) over the distribution of F (so that the term with F disappears because $\langle F\rangle=0$ by assumption), and then integrating the resulting equations for the average velocity; and the probability distribution of a velocity component v at time t is the Gaussian:

8.3.5
$$G(v,t)=\Big(\frac{m}{2\pi k_BT(1-e^{-2\beta t})}\Big)^{1/2}\exp\Big\{-\frac{m}{2k_BT}\frac{(v-\overline{v}(t))^2}{(1-e^{-2\beta t})}\Big\}\,. \tag{8.3.5}$$

---

[3] Starting from a Gaussian probability distribution over n variables $x_1,\ldots,x_n$ of the form

$$\pi(d\underline{x})=const\,e^{-\frac{1}{2}\sum_{i,j}M_{i,j}x_ix_j}\prod_h dx_h \tag{$*$}$$

where M is a positive definite symmetric matrix, then:

$$\int\pi(d\underline{x})\,x_ix_j=(M^{-1})_{i,j} \tag{$**$}$$

defining the *covariance* of the Gaussian process π and identifying it with the inverse of the matrix of the quadratic form defining π. *Also*, for all vectors $\underline{\varphi}=(\varphi_1,\ldots,\varphi_n)$, we have:

$$\int\pi(d\underline{x})\,e^{\sum_j\varphi_jx_j}=e^{\frac{1}{2}\sum_{i.j}(M^{-1})_{i,j}\varphi_i\varphi_j}\,. \tag{$***$}$$

Viceversa $**$ plus Gaussianity of π implies the other two and also $(***)$ implies the other two (and gaussianity of π). Finally one should remark that *any random variable which is a linear combination of the Gaussian variables* x_i has a probability distribution that is Gaussian. One now replaces sums with integrals and matrices with operators and one obtains the corresponding notions and relations for a Gaussian stochastic process indexed by a continuum time label.

This is computed also from (8.3.1) by writing it as

8.3.6
$$v(t) = e^{-\beta t} v_0 + \int_0^t \frac{d\tau}{m} e^{-\beta(t-\tau)} F(\tau) \tag{8.3.6}$$

so that squaring both sides and using (8.3.2) one gets

8.3.7
$$\begin{aligned}&\langle (v(t) - e^{-\beta t} v_0)^2 \rangle = \\ &= \int_0^t d\tau\, d\tau'\, e^{-\beta(t-\tau)} e^{-\beta(t-\tau')} \frac{f^2}{m^2}\, \delta(t-t') \;=\; \frac{f^2}{m^2} \frac{(1-e^{-2\beta t})}{2\beta}\end{aligned} \tag{8.3.7}$$

hence (8.3.5) follows because the distribution of $v(t)$ must be Gaussian since (8.3.6) shows that $v(t)$ is a linear combination of Gaussian variables, see footnote 3.

By integrating (8.3.6) once more one obtains in the same way also the distribution of the position component $s(t)$. It is a Gaussian with center at $\overline{s}(t)$ and quadratic dispersion

8.3.8
$$\sigma(t) = \frac{k_B T}{m\beta^2}(2\beta t - 3 + 4e^{-\beta t} - e^{-2\beta t}) \xrightarrow[t\to\infty]{} \frac{2k_B T}{m\beta} t \equiv 2D_{einst.} t \tag{8.3.8}$$

or

8.3.9
$$H(s,t) = \Big(\frac{1}{2\pi\sigma(t)}\Big)^{1/2} e^{-(s-\overline{s}(t))^2/2\sigma(t)} \tag{8.3.9}$$

The formulae just described reduce to the previous ones of Einstein's theory in the limit $t \to \infty$, but they hold *also* if $t < t_0 = \frac{m}{\lambda}$ and hence they solve the problem of the colloidal particle motions over time scales of the order of t_0 or less.

The relation $\frac{k_B T}{m\beta} = D$ (*i.e.* $k_B T \lambda^{-1} = D$) connecting viscosity (or "dissipation") and microscopic force fluctuations due to collisions with the solvent was the first example of a series of similar relations called "*fluctuation–dissipation theorems*".

Uhlenbeck and Ornstein also computed the "*joint*" probability distributions of the values $v(t_1), s(t_1), \ldots, v(t_n), s(t_n)$ for arbitrary $t_1, \ldots t_n$, hence the resulting Gaussian process (*i.e.* the probability distribution of the two component functions $t \to (v(t), s(t))$) is therefore called a Ornstein–Uhlenbeck process.

A modern discussion of the theory can be found in [Sp91], part I, Chap.8: here the motion of a Brownian particle is discussed by treating it is a "tracer" revealing the underlying microscopic motions. This is perhaps the main role of Brownian motion in macroscopic physics.

§8.4. Wiener's Theory.

From a mathematical viewpoint one can consider an idealized random motion with the property that the position $\underline{r}$ at a time $\overline{t} + t$ relative to that at

time $\bar{t}$ has, *even for an extremely small time* t, the probability given by

8.4.1
$$P(\underline{r},t)=\frac{e^{-\underline{r}^2/4Dt}}{(4\pi Dt)^{3/2}} \tag{8.4.1}$$

i.e. given, *even for small time*, by the *asymptotic* distribution (as $t\to\infty$ or $t\gg t_0$) of the Brownian motion.

Clearly the proportionality of $\underline{r}^2$ to t (rather than to t^2) for small t means that one shall find a motion with the remarkable property of *not* having a well-defined velocity at *any time* just as the Brownian motion observed over time scales longer than the previously introduced t_0.

The very possibility of rigorously *defining* such an object is the remarkable contribution of Wiener (1923), who showed that the Gaussian process with transition probability (8.4.1) (already introduced by Bachelier in the above quoted article, [Ba00]) is well defined from a mathematical viewpoint and that *with probability* 1 the paths described by the particles are *continuous*, and in fact Hölder continuous with exponent α (with any $\alpha<1/2$, see (8.4.7) below) (*Wiener theorem*, [Ne67],[IM65]).

The Gaussian process describing the probability of trajectories $t\to\underline{r}(t)$ in which the *increments* of $\underline{r}$ are distributed independently and with Gaussian distribution (8.4.1) is, in probability theory, called a *Wiener process* or, simply, a (mathematical) Brownian motion. From the Physics viewpoint it corresponds to the description of the asymptotic behavior of a colloidal particle in a fluid, for times t large compared to the characteristic relaxation time t_0 (while for generic times, including the short times, it is rather described by the Ornstein–Uhlenbeck process).

More technically we can translate into a rather simple formula the statement that the *increments* of one (of the three) coordinate $\omega(t)-\omega(t')$ of a Brownian motion are *indpendent* and distributed according to (8.4.1). This means that if $0<t_1<t_2<\ldots<t_n$ then the probability p that $\omega(t_1)\in dx_1,\omega(t_2)\in dx_2,\omega(t_n)\in dx_n$ is given by:

8.4.2
$$p=\prod_{j=1}^n\frac{dx_j}{\sqrt{4\pi D(t_j-t_{j-1})}}e^{-\frac{1}{4}\frac{(x_j-x_{j-1})^2}{D(t_j-t_{j-1})}},\qquad x_0=0,\,t_0=0\,. \tag{8.4.2}$$

Although Wiener's process is, as we have seen, a "mathematical abstraction" it has, nevertheless, great theoretical interest and it appears in the most diverse fields of Physics and Mathematics.

Its first application was to provide several *quadrature formulae* that express solution of various partial differential equations in an "*explicit form*", as integrals over families of curves randomly distributed with a Wiener process law, [Ne67].

Obviously the calculation of such integrals is, usually, *not simpler* than the solution of the same equations with more traditional methods. Nevertheless the explicit nature of the formulae provides an intuitive representation of the solutions of certain partial differential equations and often leads to surprisingly simple and strong *a priori* estimates of their properties.

A classical example is the theory of the heat equation:

8.4.3 $$\partial_t u = D\,\Delta u, \qquad u|_{t=0} = u_0(\underline{x}) \tag{8.4.3}$$

whose solution can be written as

8.4.4 $$u(\underline{x},t) = \int d\underline{y} \int P(d\underline{\omega})^t_{\underline{x},\underline{y}} u_0(\underline{y}) \tag{8.4.4}$$

where the integral is extended over all *continuous* curves $\tau \to \underline{\omega}(\tau)$ that at $\tau = 0$ start from the position $\underline{y}$ arriving at time t at $\underline{x}$. The integration ("sum") over the paths is performed by using the distribution of the Wiener process, *subject to the condition of reaching* $\underline{x}$ *at time* t.

Equation (8.4.3) has the following interpretation: *heat undergoes a Brownian motion*, *i.e.* it is transferred from point to point following a random motion with distribution given by Wiener process. Therefore the amount of heat $u(\underline{x},t)$ which at time t is in $\underline{x}$ can be obtained by imagining that the amount of heat initially in a generic point $\underline{y}$ is equitably distributed among all trajectories of the Wiener process that leave $\underline{y}$, so that the amount of heat that one finds in $\underline{x}$ at time t (*i.e.* $u(\underline{x},t)$) is the sum over all Brownian paths that arrive at $\underline{x}$ each carrying an amount of heat proportional to the amount $u_0(\underline{y})d\underline{y}$ initially around the point $\underline{y}$ where they originated; and the proportionality factor is precisely equal to the fraction of Brownian trajectories that start in the volume element $d\underline{y}$ and arrive at $\underline{x}$ in the time interval t.

Formula (8.4.4) is the simplest instance of a class of formulae that *solve* partial differential equations; a further classical example is provided by the equation

8.4.5 $$\partial_t u = D\,\Delta u + V(\underline{x})u, \qquad u|_{t=0} = u_0(\underline{x}) \tag{8.4.5}$$

which can be *explicitly* solved by the quadrature

8.4.6 $$u(\underline{x},t) = \int d\underline{y} \int P^t_{\underline{x},\underline{y}}(d\underline{\omega}) e^{-\int_0^t V(\underline{\omega}(\tau))d\tau} u_0(\underline{y}) \tag{8.4.6}$$

which is called the *Feynman–Kac quadrature formula*.

As an example of a simple application of (8.4.6) one can derive a *comparison theorem* for solutions of the equation $\partial_t u = D\Delta u + V_j(\underline{x})\,u$, $u(0) = u_0 \ge 0$ where $V_j(\underline{x})$, $j = 1,2$, are two functions (not necessarily positive) such that $V_2(\underline{x}) \le V_1(x)$. Then (8.4.6) immediately implies that $u_2(\underline{x},t) \ge u_1(\underline{x},t)$ for all $\underline{x}$ and $t \ge 0$, a property that is not so easy to prove otherwise.

Equation (8.4.6) admits various extensions, relevant both in mathematics and in physics in very diverse fields ranging from probability theory or partial differential equations to statisical mechanics and relativistic Quantum field theory and even to the foundations of quantum mechanics (see the analysis of hidden variables in Nelson's or in Böhm's quantum mechanics formulations, [Ne67], [BH93]). One can say that in these applications the formulae of explicit solution really play a role similar to that played by the classical quadrature formulae in classical mechanics.

Wiener's process had, and still has, a particular importance in probability theory, where it introduced a wealth of ideas and problems (and provided solutions to several preexisting problems). We quote the following as examples among the mathematical properties of the process.

(1) *P. Levy's regularity law.* This gives the behavior of the variation $\omega(t) - \omega(t')$ of a position component $\omega(t)$ of a Wiener process trajectory $t \to \underline{\omega}(t)$ observed at two nearby instants t, t' within a prefixed time interval $[0, \overline{t}]$. It has already been said that Wiener proves that the trajectory is Hölder continuous with an exponent α that can be fixed arbitrarily provided it is smaller than $\frac{1}{2}$. This means that, for arbitrarily fixed $\alpha < 1/2$, we shall have *with probability* 1:

8.4.7
$$\lim_{t-t'\to 0} \frac{|\omega(t) - \omega(t')|}{|t-t'|^\alpha} = 0 \tag{8.4.7}$$

if $0 \le t, t' \le \bar{t}$. The *arbitrariness* of $\alpha < 1/2$ makes it interesting to ask what is the "optimal" value for α, if any. Levy's law says that the is no optimal α, but at the same time it provides an answer to what is the actual regularity of a trajectory ω because it states

8.4.8
$$\limsup_{\substack{t-t'\to 0\\ 0\le t,t'\le \bar{t}}} \frac{|\omega(t) - \omega(t')|}{(4D|t-t'|\log\frac{1}{|t-t'|})^{1/2}} = 1 \tag{8.4.8}$$

with probability 1, [IM65].

(2) But the Levy's regularity law does not provide us with informations about the properties of the trajectory in the vicinity of a given instant: in fact (8.4.8) only gives the *worst behavior*, *i.e.* it only measures the maximal lack of regularity within a specified time interval $[0, \overline{t}]$. If we concentrate on a given instant $\tilde{t}$ then, in general, the trajectory will not be as irregular. This is in fact the content of the *iterated logarithm law* of Kintchin: the law gives the regularity property of a trajectory at a prefixed instant $\tilde{t}$. Fixing $\tilde{t} = 0$, and supposing $\omega(0) = 0$, the law is

8.4.9
$$\limsup_{t\to 0} \frac{|\omega(t)|}{\left(4Dt\log\big(\log\frac{1}{t}\big)\right)^{1/2}} = 1 \tag{8.4.9}$$

with probability 1. Equation (8.4.9) is not incompatible with (8.4.8). In fact it only says that the worst possible behavior described by (8.4.8) is in fact not true with probability 1 at a prefixed instant *i.e. it happens certainly*, by the previous law, but *certainly* as well *it does not happen* at the time $\tilde{t}$ at which one has decided to look at the motion! [IM65].

(3) The above two laws deal with the behavior of the trajectories at *finite times*; one can ask what is the long-time behavior of a sampled trajectory. The Einstein and Smoluchowski theories foresee that the motion goes away

from the vicinity of the origin (*i.e.* of the starting point) by a distance that grows proportionally to $\sqrt{t}$.

An analysis of these theories indicates that by *average* one means average over a statistical ensemble. If at time t one measures the square of a coordinate of the particle which at time 0 was at the origin and if one does this for many Brownian particles (*i.e.* if one repeats the measurement many times) one finds $2Dt$, in the average.

However this does not mean that if one fixes attention on a *single* motion and one observes it as t grows, every coordinate $\omega(t)$ squared grows at most as $2Dt$ in the sense that the maximum limit as $t \to \infty$ of $\omega(t)^2/t$ is $2D$. In fact such growth is really given by the *global iterated logarithm law* (Kintchin):

8.4.10
$$\limsup_{t\to+\infty} \frac{|\omega(t)|}{\big(4Dt\log(\log t)\big)^{1/2}} = 1 \tag{8.4.10}$$

with probability 1, [IM65].

Although (8.4.10) deals with Wiener process properties it expresses a property of relevance for experiments on Brownian motion: *unlike* the two previous laws, which look at properties characteristic of the Wiener process and not of real Brownian motions (which are rather described by the Ornstein–Uhlenbeck process), this is a property that refers to the large time behavior (which is the *same* for the Wiener process and for the Ornstein–Uhlenbeck process). However it is very difficult to perform experiments so accurate as to reveal a correction to the displacement which is proportional to the square root of an iterated logarithm.

(4) Equation (8.4.10) does not invalidate the measurability of D based on the observation of a single trajectory. Such measurements are performed by following the displacement $\omega(t)$ of a coordinate as t varies between 0 and t. One then sets

8.4.11
$$X(t) = t^{-1}\int_0^t \omega(\tau)^2 d\tau \tag{8.4.11}$$

and a fit is attempted by comparing the data $X(t)$ with the function $2Dt$. The procedure is correct, at least asymptotically as $t \to \infty$, because one shows that:

8.4.12
$$\lim_{t\to\infty} \frac{X(t)}{2Dt} = 1 \tag{8.4.12}$$

with probability 1. This is the *ergodic theorem* for the Wiener process. Like the comparison between laws (1), (2) above, law (3) tells us that our particle will be "far" ahead of where it should infinitely many times (*i.e.* a factor $(\log\log t)^{\frac{1}{2}}$ ahead, (8.4.10)) although in the average it will be at a distance proportional to t ((8.4.12)), [IM65].

(5) The trajectories run by the Wiener process are rather *irregular*, as the Levy and Kintchin laws quantitatively show. One can ask which is the *fractal dimension* of the set described by a Wiener process trajectory. If

the dimension of the space in which the motion take place is ≥ 2 then the fractal dimension is 2, in the sense of Hausdorff. This essentially means that if one wants to cover the trajectory with spheres of radius $1/n$ one needs $O(n^2)$ spheres in the sense that given $\varepsilon > 0$ then $O(n^{2-\varepsilon})$ are (eventually as $n \to \infty$) not sufficient while $O(n^{2+\varepsilon})$ are sufficient (even if $n \to \infty$). Note that a smooth curve, with finite length, can be covered by just $O(n)$ spheres of radius n^{-1}.

The property of showing dimension 2 can be expressed also in several other ways which are perhaps intuitively equivalent, at a superficial level of understanding, but strictly speaking different, and the analysis of the alternative ways illustrates subtle aspects of the Wiener's process trajectories.

For instance if one considers two distinct points in R^d and from each of them one starts a Wiener path, then one finds that the two paths will eventually "cross" (*i.e.* they reach the same point at some later time) with probability 1 *if* $d = 2, 3$, as expected intuitively on the basis that dimensionally they are "surfaces". *But* they *do not* cross if $d \geq 4$ (with probability 1): if they were really 2–dimensional geometric objects we would expect that they would intersect not only for $d = 2, 3$ but for $d = 4$ as well. This is *Lawler's theorem*, [La85], see also [HS92].

(6) A further celebrated property of the Wiener process, due to Wiener himself, exhibits interesting connections with harmonic analysis and Fourier series theory. Consider a sequence $g_0, g_1, \ldots$ of Gaussian independent and equidistributed random variables and suppose that the distribution of each of them is $(2\pi)^{-1/2} \exp -g^2/2$. Set

8.4.13
$$\omega(t) = \frac{t}{\pi^{1/2}} g_0 + \left(\frac{2}{\pi}\right)^{1/2} \sum_{k \geq 1} \frac{\sin kt}{k}\, g_k\,. \tag{8.4.13}$$

Then the random function $\omega(t)$, for $0 \leq t \leq \pi$, has a probability distribution (induced by the one assumed for the coefficients g_k) *identical* to that of a Wiener process sample path in dimension 1 (and $D = \frac{1}{2}$), [IM65]. This remarkable fact is easy to check; since the covariance of a Gaussian process determines the process, it suffices to check that if $t > t'$ then, see (8.4.2):

8.4.14
$$\langle \omega(t)\omega(t') \rangle = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \int e^{-\frac{1}{2}\frac{x^2}{t}} e^{-\frac{1}{2}\frac{(x-y)^2}{(t-t')}}\, x\, y\, dx\, dy \tag{8.4.14}$$

The left-hand side is immediately computable from (8.4.13) and from the assumed Gaussian distribution of the g_k's and the r.h.s. is an elementary integral so that the identity is easily checked.

To conclude one can say that the Wiener process is a mathematical abstraction originated from the physical phenomenon of Brownian motion: it describes its "large time" behavior (the behavior for all times being caught, more appropriately, by the Ornstein–Uhlenbeck process). It is nevertheless a mathematical entity of great interest which finds applications in the most different (and unexpected) fields of mathematics and physics. Further reading on both the mathematical and physical aspects of Brownian motion can be found in [Sp91].

Chapter IX

Coarse Graining and Nonequilibrium

.

§9.1. Ergodic Hypothesis Revisited

An informal overview of the basic ideas of Chap.IX is given in Appendix 9A1 below, which is the text of a conference at the Séminaire de Philosophie et Mathématiques of École Normale Superieure in Paris.

Giving up a detailed description of microscopic motion led to a statistical theory of macroscopic systems and to a deep understanding of their equilibrium properties which we have discussed in Chap.I-VII.

It is clear today, as it was already to Boltzmann and many others, that some of the assumptions and guiding ideas used in building up the theory were not really necessary or, at least, could be greatly weakened or just avoided.

A typical example is the *ergodic hypothesis*. Although we have analyzed it in some detail in Chap.I it is interesting to revisit it from a different perspective. The analysis will not only help clarify aspects of nonequilibrium statistical mechanics, but it will be important for its very foundations.

We have seen the important role played by the *heat theorem* of Boltzmann, [Bo84]. We recall that one can define in terms of time averages of total or kinetic energy, of density, and of average momentum transfer to the container walls, quantities that one could call, respectively, *specific internal energy* u, *temperature* T, *specific volume* v, *pressure* p; and the heat theorem states that when two of them varied, say the specific energy and volume by du and dv, the relation

9.1.1
$$\frac{du + p\,dv}{T} = \text{ exact} \tag{9.1.1}$$

holds.

In the beginning, [Bo66], this was discussed in very special cases (like free gases), but about fifteen years later Helmholtz, influenced by the progress of Boltzmann on the proof of the heat theorem, wrote a series of four ponderous papers on a class of very special systems, which he called *monocyclic*, in which all motions were periodic and in a sense non-degenerate, and he noted that one could give appropriate names, familiar in macroscopic thermodynamics, to various mechanical averages and then check that they satisfied the relations that would be expected between the thermodynamic quantities with the same name.

Helmholtz' assumptions about monocyclicity are very strong and seem to be satisfied in no system other than in confined one-dimensional Hamiltonian systems. Here are the details of Helmholtz' reasoning (as reported by Boltzmann), in a simple example.

Consider a one-dimensional system in a confining potential.[1] There is only one motion per energy value (up to a shift of the initial datum along its trajectory) and all motions are periodic so that the system is *monocyclic*. We suppose that the potential $\varphi(x)$ depends on a parameter V.

---

[1] A potential $\varphi(x)$ such that $|\varphi'(x)| > 0$ for $|x| > 0$, $\varphi''(0) > 0$ and $\varphi(x) \xrightarrow[x\to\infty]{} +\infty$.

Suppose that one identifies a *state* with a motion with given energy E and given V. Then, let

U = total energy of the system $\equiv K+\varphi$,
T = time average of the kinetic energy K,
V = the parameter on which φ is supposed to depend,
p = $-$ time average of $\partial_V\varphi$.

A state is parameterized by U,V and if such parameters change by dU,dV respectively we define:

9.1.2
$$dL=-pdV,\qquad dQ=dU-dL\,. \tag{9.1.2}$$

then:

Theorem (Helmholtz): *The differential $dQ/T=(dU+p\,dV)/T$ is exact.*

Repeating, for convenience, the few lines of proof already discussed in Appendix 1.A1 to Chap.I, this can be proved by directly exhibiting a function S whose differential is $(dU+pdV)/T$. In fact let $x_\pm(U,V)$ be the extremes of the oscillations of the motion with given U,V and define S as

9.1.3
$$S(U,V)=2\log\int_{x_-(U,V)}^{x_+(U,V)}\sqrt{U-\varphi(x)}\,dx \tag{9.1.3}$$

and $\frac12 S$ is the logarithm of the action since $U-\varphi(x)$ is the kinetic energy $K(x;U,V)$; so that

9.1.4
$$dS=\frac{\int(dU-\partial_V\varphi(x)dV)\frac{dx}{\sqrt{K}}}{\int K\frac{dx}{\sqrt{K}}}\,. \tag{9.1.4}$$

Noting that $\frac{dx}{\sqrt{K}}=\sqrt{\frac{2}{m}}dt$, we see that the time averages are obtained by integrating with respect to $\frac{dx}{\sqrt{K}}$ and dividing by the integral of $\frac{1}{\sqrt{K}}$. Hence:

9.1.5
$$dS=\frac{dU+p\,dV}{T}\,, \tag{9.1.5}$$

completing Helmholtz' remark. For a more extended discussion of the theorem see Appendix 1.A1 to Chap.I.

Boltzmann saw that *this was not a simple coincidence*: his interesting (and healthy) view of the continuum which, probably, he never really considered more than a convenient artifact, useful for computing quantities describing a discrete world where sums and differences could be approximated by integrals and derivatives, cf. §1.9 and [Bo74] p. 43, led him to think that in some sense *monocyclicity was not a strong assumption.*

Motions tend to recur (and they do in systems with a discrete phase space) and in this light monocyclicity would simply mean that, waiting

long enough, the system would come back to its initial state. Thus its motion would be monocyclic and one could try to apply Helmholtz' ideas (in turn based on his own previous work) and perhaps deduce the heat theorem in great generality. The nondegeneracy of monocyclic systems becomes the condition that for each energy there is just one cycle and *the motion visits successively all* (discrete) *phase space points*.

Taking this viewpoint one had the *possibility* of checking that in all mechanical systems one could define quantities that one could name with "thermodynamic names" and which would satisfy properties coinciding with those that thermodynamics would predict for them, see Chap.I, II.

He then considered the two-body problem, showing that the thermodynamic analogies of Helmholtz could be extended to systems which were degenerate, but still with all motions periodic. This led to somewhat obscure considerations that seemed to play an important role for him, given the importance he gave them. They certainly do not help in encouraging reading his work: the breakthrough paper of 1884, [Bo84], starts with associating quantities with a thermodynamic name to Saturn's rings (regarded as rigid rotating rings!) and checking that they satisfy the right relations, like the second principle, see (9.1.1).

In general one can call *monocyclic* a system with the property that there is a curve $\ell \to x(\ell)$, parameterized by its curvilinear abscissa ℓ, varying in an interval $0 < \ell < L(E)$, closed and such that $x(\ell)$ covers all the positions compatible with the given energy E.

Let $x = x(\ell)$ be the parametric equations so that energy conservation can be written, for some $m > 0$,

9.1.6 $$\frac{1}{2} m\,\dot\ell^2 + \varphi(x(\ell)) = E\,. \tag{9.1.6}$$

then if we suppose that the potential energy φ depends on a parameter V and if T is the average kinetic energy, $p = -\langle\partial_V\varphi\rangle$ then, for some S,

9.1.7 $$dS = \frac{dE + p\,dV}{T}, \qquad p = -\langle\partial_V\varphi\rangle, \quad T = \langle K\rangle \tag{9.1.7}$$

where $\langle\cdot\rangle$ denotes the time average (see Appendix 1.A1, Chap.I).

A typical case to which the above can be applied is the case in which the whole energy surface consists of just one periodic orbit, or when at least only the phase space points that are on such orbit are observable. Such systems provide, therefore, natural models of thermodynamic behavior.

A chaotic system like a gas in a container of volume V, which will be regarded as an important parameter on which the potential φ (which *includes* interaction with the container walls) depends, will satisfy "for practical purposes" the above property, because (Feynman) "*if we follow our solution* [*i.e.* motion] *for a long enough time it tries everything that it can do, so to speak*" (see p. 46-55 in [Fe63], vol. I). Hence we see that we should be able to find a quantity p such that $dE + p\,dV$ admits the average kinetic energy as an integrating factor.

On the other hand if we accept the viewpoint (ergodic hypothesis) that phase space is discrete and motion on the energy surface is a monocyclic permutation of its finitely many cells, the *time averages can be computed by integrals* with respect to the uniform distribution, that we shall call *Liouville distribution*,[2] see §1.7.

Hence if μ is the Liouville's distribution on the surface of constant energy U, and T is the μ-average kinetic energy then there should exist a function p such that T^{-1} is the integrating factor of $dU + pdV$.

Boltzmann shows that this is the case and, in fact, p is the μ-average $\langle -\partial_V \varphi \rangle$ and it is *also* the average momentum transfer to the walls per unit time and unit surface, *i.e.* it is the *physical* pressure, see Appendix 9.A3.

This *is not a proof* that the equilibria are described by distributions μ of the the microcanonical ensemble. However it shows that for most systems, independently of the number of degrees of freedom, one can define a *mechanical model of thermodynamics*, *i.e.* one can define various averages of mechanical quantities and name them with names of thermodynamical functions, and check that they satisfy the relations that would follow from classical thermodynamics, see Chap.I,II for more details.

thermodynamic relations are, therefore, very general and simple consequences of the structure of the equations of motion. They hold for small and large systems, from one degree of freedom (the case of Helmoltz' monocycles) to 10^{23} *degrees and more (the case of a gas in a box).*

The above arguments, based on a discrete view of phase space, suggest however that, in general, the thermodynamic relations hold in some approximate sense, as we have no idea of the precise nature of the discrete phase space. However, in some cases, they may hold *exactly even for small systems*, if suitably reformulated: for instance in the 1884 paper, [Bo84], Boltzmann shows that in the *canonical ensemble* the relation (9.1.1) (*i.e.* the second law) holds *without corrections* even if the system is small, as explained in Chap.II.

Thus the ergodic hypothesis does help in finding out *why* there are mechanical "models" of thermodynamics: they are ubiquitous, in small and large systems alike, but usually such relations are of interest in large systems and not really in small ones.

A critical comment and a warning is important at this point: for large systems *any theory claiming to rest on the ergodic hypothesis may seem bound to fail*, see §1.7, because if it is true that a system is ergodic, it is also true that the time the system takes to go through one of its cycles is simply too long to be of any interest and relevance: this was pointed out very clearly by Boltzmann, [Bo96], and earlier by Thomson, [Th74].

The reason why we observe approach to equilibrium over time scales far shorter than the recurrence times is due to the property that the micro-

---

[2] Which is the only invariant distribution if one accepts the above discrete point of view, probably Boltzmann's.

canonical ensemble is such that *on most of phase space the actual values of the observables, whose averages yield the pressure and temperature and the few remaining other thermodynamic quantities, assume the same value*, [La72], and p. 206 of [Bo74]. This implies that such values coincide with the average and *therefore satisfy the heat theorem.*

The ergodic hypothesis loses its importance and fundamental nature and it appears simply as a tool used in understanding that some of the relations that we call "macroscopic laws" hold in the same form for all systems, whether small or large.

§**9.2. Timed Observations and Discrete Time**

The question that we shall now investigate is whether there can be anything similar to the above done out of equilibrium but still in a stationary state: are there statistical properties that *hold for small and large systems alike* under the "only" assumption that the systems evolve in a very disordered, "chaotic", way? If so such properties might have physical relevance for large systems: when the system is large they may become observable because they may become a property of most of the individual configurations of the system without need of time averaging, just as it happens in equilibrium.

And they might be checked, and perhaps even be interesting, in small systems which become therefore a natural testing ground, mainly because of the availability of fast computer experiments, just as it happens in equilibrium with the ergodic hypothesis, which is usually tested in systems with very few degrees of freedom.

The first step in the investigation is a dynamical hypothesis on the nature of the motions of complex systems (like a gas in a box). This hypothesis has developed quite slowly in the past epoch: it developed from the theory of complex motions in fluid mechanics and it was formulated by Ruelle in the early 1970s (1973) and written explicitly later, [Ru80], [Ru76]. It influenced research strongly, see for instance [ECM90],[ECM93]; and it led to some concrete results, after being reformulated and put in the context of nonequilibrium statistical mechanics, "much later" [GC95]. The hypothesis will be stated below and is called the *chaotic hypothesis.*

To proceed to the formulation of the hypothesis we need to set up a convenient kinematic description of disordered motions, convenient for the study of chaotic evolutions: this is necessary because the usual kinematics is well suited for orderly motions but is insufficient for disordered ones.

In Chap.I we have already hit the difficulty of a proper representation of the evolution of a system of N particles in a box V as a permutation of phase space cells. The difficulty came from the *hyperbolic* nature of the evolution that stretches some coordinates and contracts others. This forced us to use very small cells and very small time intervals as phase space and time units.

The division in cells of phase space is, in this way, extremely fine (although, as we discussed, there are conceptual limits to the precision and reliability that one can reach in this way), so fine that we can regard space as discrete.

It is interesting to change viewpoint and, keeping the conception of space as a continuum, to try using a discrete representation of motion based on phase space cells that *are not as small as possible* (but that are nevertheless still compatible with the discrete picture that we follow after Chap.I.): we shall call such a description a "*coarse grained*" one.

In other words it may be more convenient to use larger phase space cells and to find a description of dynamics in terms of them, which explicitly and "exactly" still takes into account the (often) "hyperbolic" nature of the evolution, *i.e.* its (usual) high instability or its *chaoticity*.

The use of the "coarse grained" cells brings to mind something "approximate" and not too well defined. Here we *do not want* to convey such intuition and the representation of motion, that we look for, will be in principle as exact as wished, and involve no approximation at all. The modern efforts to clarify the notion of coarse graining can be traced back to Krylov, [Kr79], whose work strongly influenced Sinai, [Si79], whose work in turn influenced Ruelle leading him, eventually, to his hypothesis.

We consider a system evolving on a bounded surface Σ and with the evolution acting near a given point x by expanding some line elements and by contracting some others. The evolution will be described by *a map* S that can be thought of as being obtained from the time evolution flow $x \to S_t x$ by monitoring it every time some special event happens (for instance a collision between some pair of particles). In this way the surface Σ consists of the collection of the special events that one monitors, which will be called *timing events* or *monitored events*.

The geometrical meaning of the construction of S from the flow S_t is illustrated in Fig. 9.2.1 where a trajectory $x \to S_t x$ in the phase space for the evolution in continuous time ("usual phase space") is depicted together with the surface Σ consisting of the *monitored events* (which is the phase space where the motion is described by the map S: the map S acts, in the Fig. 9.2.1, on the monitored event ξ mapping it into $S\xi$ and $S^2\xi$).

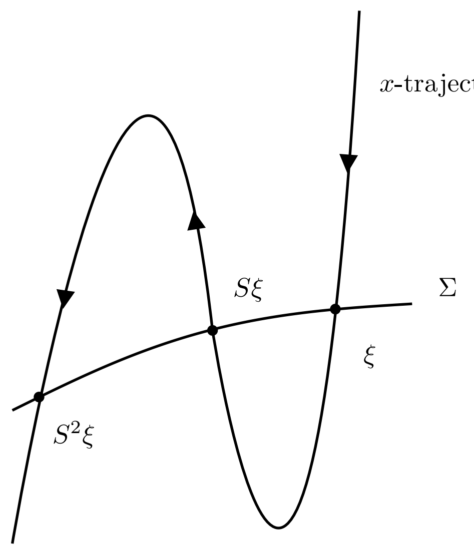


Fig. 9.2.1

Line elements on Σ are transformed by the time evolution S and some of them become longer and others shorter (or keep their length). Note that in order to measure the length of a line element we need a metric defined on Σ: we shall take the natural metric induced on the surface Σ by the metric in the space in which Σ lies (which is usually the Euclidean space of position and velocity vectors of the system particles). But of course the metric that we use is *rather arbitrary* and we have to be careful, it is best to try dealing only with notions that turn out to be metric independent.

In the following we shall always use the above discrete time evolution, but all that we say can be quite easily translated in terms of properties of the continuous time flow S_t.[3]

§9.3. Chaotic Hypothesis. Anosov Systems

Boltzmann's equation and Brownian motion theory are examples of attempts at studying nonequilibrium problems. The first has the ambition of discussing the approach to equilibrium, while the second deals with motions that take place in equilibrium.

In general we shall say that a system is in a *nonequilibrium* situation whenever *nonconservative* external forces act on it and, usually, sustain macroscopic motions. Such systems will evolve and reach in due course a *stationary state*, which *will not be one of the equilibrium states* with which we are familiar from the preceding chapters.

This is because time evolution will ultimately be described by differential

[3] One should not confuse this map with the map used in Chap.I for the one time step evolution. Two successive timing events still contain very many time steps in the sense of Chap.I: here we do not discretize time or space. Motion appears evolving discretely simply because we *choose* to observe it from time to time, when something that we consider interesting happens.

equations which will contain dissipative terms and consequently phase space volume will be contracting; hence probability distributions corresponding to the statistics of stationary states *must* be concentrated on sets with *zero volume*.

Before attempting a general theory of the approach to an equilibrium state or, more generally, to a stationary state it seems reasonable to study the properties of the stationary states themselves, end products of the evolution under external driving forces (which may vanish, however, so that equilibrium theory will still be a "special case").

This will include investigating phenomena like Brownian motion which can be regarded as dynamical properties of equilibrium states, as well as genuinely nonequilibrium phenomena, like thermodynamical relations between quantities that can be defined in systems in nonequilibrium stationary states.

And one can imagine stretching the analysis to stationary states of the macroscopic equations that are supposed to be obeyed out of equilibrium, like the Euler or Navier-Stokes equations for systems macroscopically behaving as fluids.

Approach to equilibrium or to a stationary state is likely to be a more difficult problem and it will be set aside in most of what follows. Hence we shall not deal with states of systems evolving in time: rather we refer to properties of states that *are already* in a stationary state under the influence of external nonconservative forces acting on them. For instance think of an electric circuit in which a current flows (stationarily) under the influence of an electromotive field, or of a metal bar with two different temperatures fixed at the extremes; and one can even think of a Navier-Stokes fluid in a (stationary) turbulent Couette flow or a more general flow. What follows applies also to such apparently different systems (and in fact the basic ideas were developed having precisely such systems in mind).

The first two systems, regarded as microscopic systems (*i.e.* as mechanical systems of particles), do certainly have very chaotic microscopic motions even in the absence of external driving (while macroscopically they are in a stationary state and nothing happens, besides a continuous, sometimes desired, heat transfer from the system to the surroundings). The third system also behaves, as a macroscopic system, very chaotically at least when the Reynolds number is large.

A basic problem is that the situation is quite different from that in which Boltzmann was when attempting a microscopic proof of the heat theorem: there is no established nonequilibrium thermodynamics to guide us.

The great progress of the theory of stationary nonequilibrium that took place in the past century (the XX-th), at least that which was unanimously recognized as such, *only* concern properties of *incipient* nonequilibrium: *i.e.* transport properties at vanishing external fields (we think here of *Onsager's reciprocity* and of its more quantitative form given by the *Green-Kubo transport theory*). So it is by no means clear that there is any general nonequilibrium thermodynamics.

Nevertheless in 1973 a first suggestion that a general theory might be possible for nonequilibrium systems in stationary and chaotic states was made by Ruelle, eventually written down only later in [Ru78c], [Ru80]. This suggestion was made with full understanding of its ambition:

"*If one is optimistic, one may hope that the asymptotic measures will play for dissipative systems the sort of role which the Gibbs ensembles played for statistical mechanics. Even if that is the case, the difficulties encountered in statistical mechanics in going from Gibbs ensembles to a theory of phase transitions may serve as a warning that we are, for dissipative systems, not yet close to a real theory of turbulence*", [Ru78c].

The proposal is very ambitious because it suggests a *general and essentially unrestricted* answer to what should be the ensemble that describes stationary states of a system, *whether in equilibrium or not*. In the recent formulations of Cohen and Gallavotti it reads:

Chaotic hypothesis: for the purpose of studying macroscopic properties, the time evolution map S of a many-particle system can be regarded as a mixing Anosov map.[4]

We defer discussing in detail the technical notion of "mixing Anosov map" to the coming sections and the hypothesis is written here only for concreteness and later reference. For the moment it will suffice to say that mixing Anosov maps are the *paradigm of chaotic motions*: they are well understood dynamical systems which show chaotic behavior in the "purest" possible way. They play a role in nonlinear dynamics very similar to that played by harmonic oscillators in the theory of stable motions.

Remark: If the evolution is very dissipative and motions tend to an attracting set smaller than the whole phase space the hypothesis may be interpreted as meaning that the attracting set can be regarded as a smooth surface and that the restriction of the evolution to it is a mixing Anosov map, see below (which is a *very special* case of a wider class of chaotic systems called *Axiom A systems*), [BGG97],[BG97]. However a less strict interpretation could be to say that the attractor is an "Axiom A attractor", see below.

The ergodic hypothesis led Boltzmann to the general theory of ensembles (as acknowledged by Gibbs, p. vi in [Gi81], whose work has been perhaps the main channel through which the allegedly obscure works of Boltzmann reached us): besides giving the second law, (9.1.1), it also prescribed the microcanonical ensemble for describing equilibrium statistics.

The reasoning of Ruelle was that from the theory of simple chaotic systems one knew that such systems, *just by the fact that they are chaotic*, will reach

[4] This is a notion that in the original work [GC95] was called "transitive Anosov map", however it turns out that the established nomenclature is different and here I try to adhere to it, as much as possible.

a "unique" stationary state. Therefore simply assuming chaoticity would be tantamount to assuming that there is a uniquely defined ensemble which should be used to compute the statistical properties of a stationary system out of equilibrium.

This argument is based on the idea that a chaotic system, even when it is not exactly a mixing Anosov map in the mathematical sense, does share its main qualitative features, [Ru78c], in a sense similar to the one in which in stability theory one infers properties of nonlinear oscillations from those of harmonic oscillators or of integrable systems. And a key property of mixing Anosov maps is that their motions show a *unique* statistics μ. This means that there is a unique probability distribution μ on phase space $\mathcal{F}$ such that for all (smooth) observables $F(x)$:

9.3.1
$$\lim_{T\to\infty}\frac{1}{T}\sum_{n=0}^{T-1}F(S^n x)=\int_{\mathcal{F}}F(y)\,\mu(dy) \qquad (9.3.1)$$

apart from a set of zero volume of initial data $x\in\mathcal{F}$.[5]

The distribution μ is called the *SRB distribution* or the *statistics of motions*: it was proven to exist by Sinai for Anosov systems and the result was extended to the much more general Axiom A attractors by Ruelle and Bowen, [BR75], [Ru76]. *Natural distributions* were, independently, discussed and shown to exist, [LY73], for other (related and simpler) dynamical systems, although in an apparently less general context and with a less general vision of the matter, [Si70],[BR75].

Therefore one is, at least in a very theoretical way, in a position to inquire whether such a unique ensemble has *universal properties* valid for small and large systems alike (of course we cannot expect too many of them to hold, but *even a single one* would be interesting).

In fact in equilibrium theory the only universal property is precisely the heat theorem, besides a few general (related) inequalities (*e.g. positivity of the specific heat or of compressibility*). The theorem leads, indirectly as we have seen, to the microcanonical ensemble and then, after one century of work, to a rather satisfactory theory of phenomena like phase transitions, phase coexistence and universality.

From this point of view the criticized (and more and more often dismissed as, at best, unnecessary) ergodic hypothesis assumes a new status and emerges as greatly enhanced. Ruelle's proposal seems to be its natural (and, perhaps, the unique) *extension out of equilibrium.*

The proposal was formulated in the case of fluid mechanics: but it is so clearly more general that the reason why it was not explicitly referring to statistical systems is, probably, due to the fact that, as a principle, it required some "check" if formulated for statistical mechanics. As originally

---

[5] In general replacing S by S^{-1} in (9.3.1) leads, when the new limit exists, to a *different* probability distribution on phase space, which we can call the *statistics towards the past*, while (9.3.1) defines the *statistics towards the future*. One *cannot* expect that, when they exist, the two statistics coincide.

stated, and without any further check, it would have been analogous to the ergodic hypothesis without the heat theorem (or other consequences drawn from the theory of statistical ensembles).

Of course the chaotic hypothesis will suffer from the same objections that are continuously raised about the ergodic hypothesis: namely "there is the time scale problem", see §1.7.

To such objections the answer given by Boltzmann should apply *unchanged*: large systems have the extra property that the interesting observables take the same value in the whole (or virtually whole) phase space. *Therefore their values satisfy any relation that is true no matter whether the system is large or small*: such relations (whose very existence is, in fact, surprising) might even be of no interest whatsoever in small systems (like in the above-mentioned Boltzmann's rigid Saturn ring, or in his other similar example of the Moon regarded as a rigid ring rotating about the Earth).

Evidence for the nontrivial applicability of the hypothesis built up and it was repeatedly hinted at in various papers dealing with numerical experiments, mostly on very few particle systems (< 100 to give an indication), [HHP87]. In attempting at understanding one such experiment, [ECM93], the above "formal" interpretation of Ruelle's principle was formulated, [GC95], for statistical mechanics (as well as for fluid mechanics, replacing "many-particles system" with "turbulent fluid").

The hypothesis was made first in the context of reversible systems: they were in fact the subject of much of the experimental work that bloomed once the importance and relevance of reversibilty was strongly stressed, and supported by experiments, by Hoover and coworkers, for highlights see [HHP87],[EM90],[ECM90],[DPH96]. Note that saying that reversibility can be relevant to, and even facilitate, the analysis of dissipative motions is highly nontrivial and it *required insight and intellectual courage to be introduced*.

The strict interpretation of the chaotic hypothesis given in the above remark *rules out*, when the attractive set is smaller than phase space, attractors[6] with a fractal closure (*i.e.* attracting sets which are *not* smooth

---

[6] It is important to distinguish between *attracting set* and *attractor*. The first is a closed set such that all points close enough to it evolve in time keeping a distance from it that tends to 0 as $t \to +\infty$, and furthermore are "minimal" in the sense that they do not contain subsets with the same properties. Consider, to define an *attractor*, an attracting set which admits a statistics μ given by (9.3.1) for all but a zero volume set of nearby points x. Any subset C of such an attracting set with $\mu(C) = 1$ is called an attractor. More generally we can imagine choosing initial data x near an attracting set with a probability distribution μ_0 that can even be concentrated on sets of zero volume, *i.e.* that is completely different from the volume measure. Supposing that all data x but a set of zero μ_0 measure satisfy (9.3.1) (of course, in general, with a statistics μ' different from μ and μ_0-dependent). The subsets of the attracting set that have probability 1 with respect to the statistics generated by such initial data x, will be called "attractors for the data with distribution μ_0", and *they may have* 0 *measure with respect to the statistics* μ defined by (9.3.1). We see that in general the notion of attracting set is uniquely determined by the dynamics, while the notion of attractor depends also on which initial data we are willing to consider; and even once the class of initial data is chosen the notion of attractor is not uniquely defined as we can always take out of an

surfaces); to include them, one should replace, in the formulation of the chaotic hypothesis, the word "Anosov" into "Axiom A" (a similar but much weaker notion): but it seems appropriate to wait and see if there is real need of such an extension. It is certainly an *essential extension* for small systems, but it is not clear to me how relevant could fractality be when the system has 10^{23} particles). Therefore we shall concentrate our attention mainly on Anosov systems.

Finally I add one more comment on the words at the beginning of the chaotic hypothesis "for the purpose of ...": an easy critique could be that this is "vague" since it is obvious that virtually none of the systems of interest in statistical mechanics (or fluid dynamics) are Anosov systems in the mathematical sense (they are often not smooth, or obviously not hyperbolic and often not even ergodic).

Nevertheless I think that the hypothesis is well founded and it has an illustrious predecessor in the early ergodic hypothesis of Boltzmann: he was trying to prove the heat theorem; he needed the condition that the motions were periodic; he said that if they were not they could still be considered so for practical purposes "because a nonperiodic orbit can be regarded as periodic with infinite period".

Assuming that the motions were periodic (*i.e.* the systems were "monocyclic") led him to discover a hitherto unknown property of periodic motions (the heat theorem). Indeed things "went as if the motions were periodic"!, [Bo66], a first rough formulation of the ergodic hypothesis. We have discussed at length the interpretation and the interest of the ergodic hypothesis in Chap.I,II,III, and we have seen that it leads to important relations between averages; one can think that "all systems are Anosov" in the same sense.

Mathematically what is being said is that there might be general properties of Anosov systems that *might have been missed* in spite of the *vast* reasearch on the subject.

§9.4. Kinematics of Chaotic Motions. Anosov Systems

To proceed we need a more precise formulation of the notion of Anosov systems and an analysis of the kinematics of their motions. We give here an informal definition and in the next section we give a detailed discussion of the kinematics of motions.

Definition: A mixing Anosov system is, see p.55 of [AA68], a smooth map S (i.e. of class C^∞) of a smooth manifold M ("phase space") and around every point x one can set up a local coordinate system with the following properties associated with it:

(a) depends continuously on x and is covariant (i.e. it follows x in its evolution) and

---

attractor one orbit and still have an attractor (unless the attractor consists of finitely many points).

(b) is hyperbolic, i.e. transversally to the phase space velocity of any chosen point x the motion of nearby points looks, when seen from the coordinate frame covariant with x, like a hyperbolic motion near a fixed point. This means that in a sphere of small radius δ around x there will be a connected "local stable coordinate surface", the stable local manifold $W_x^{s,\delta}$ through x, whose points have trajectories that get close to the trajectory of x at exponential speed as the time tends to $+\infty$, and a "local unstable coordinate surface", the local unstable manifold $W_x^{u,\delta}$, whose points have trajectories that get close to the trajectory of x at exponential speed as the time tends to $-\infty$. Furthermore the exponential speed of approach must admit a bound independent on x, i.e. it has to be "uniform".

(c) the global stable manifold and the global unstable manifold of every point, i.e. the sets $\cup_{n=0}^{\infty} S^n\, W_x^{u,\delta}$ and $\cup_{n=0}^{\infty} S^{-n}\, W_x^{s,\delta}$ are dense in M. This excludes the possibility that the phase space consists of disconnected invariant parts. It also excludes the case that it consists of n disconnected parts $M_1, \ldots, M_n$ cyclically permuted by S so that S^n is a mixing Anosov map as a map of M_j into itself, clearly a case in which one is "improperly defining" phase space and evolution map (which should be rather defined as M_1, say, and S^n respectively).

If only (a),(b) hold the system is simply called an "*Anosov system*". It is a theorem (Anosov) that the planes tangent to W_x^u, W_x^s are quite smoothly dependent on x; they are Hölder continuous in x (in general not more, however, because in general they are not differentiable in x even if the map S is analytic), [AA68],[Ru89].

If the system is described in continuous time the direction parallel to the velocity has to be regarded as an "extra" *neutral* direction where, on the average, no expansion nor contraction occurs. However here we shall adhere to the discrete viewpoint based on timed observations, see §9.1. For a discussion of the continuous time point of view see [BR75], [Bo74], [Ge98].

The simple but surprising and deep properties of Anosov maps are by and large very well understood, [Ru89]. Unfortunately they are not as well known among physicists as they should be: many seem confused by the language in which the above concepts are usually presented; however it is a fact that such remarkable mathematical objects (*i.e.* Anosov systems) have been introduced by mathematicians, and physicists must, therefore, make an effort at understanding the new notion and its physical significance.

In particular, as mentioned above, if a system is Anosov, *for all* observables F (*i.e.* continuous functions on phase space) and for all initial data x, outside a set of zero volume, the time average of F exists and can be computed by a phase space integral with respect to a distribution μ uniquely determined on phase space $\mathcal{F}$ as expressed by (9.3.1).

Clearly the chaotic hypothesis solves *in general* (*i.e.* for systems that can be regarded as "chaotic") the problem of determining which is the ensemble to use to study the statistics of stationary systems in or out of equilibrium

(it clearly implies the ergodic hypothesis in equilibrium), in the same sense in which the ergodic hypothesis solves the equilibrium case.

The chaotic hypothesis might turn out to be false in interesting cases; like the ergodic hypothesis which does not hold for the simplest systems studied in statistical mechanics, like the free gas, the harmonic chain and black body radiation. Worse, it is *known* to be false for trivial reasons in some systems in equilibrium (like the hard core gas), simply because the Anosov map definition requires smoothness of the evolution and systems with collisions are not smooth systems (in the sense that the trajectories are not differentiable as functions of the initial data).

However, interestingly enough, the hard core systems are perhaps the object *closest* to an Anosov system that can be thought of, being at the same time of statistical mechanical relevance, [GG94], [ACG96], to the extent that there seem to be no known "physical" properties that this system does not share with an Anosov system. Aside from the trivial fact that it is not a smooth system, the hard core system behaves, for statistical mechanics purposes, *as if it was a (mixing) Anosov system.* Hence it is the prototype system to study in looking for applications of the chaotic hypothesis.

In fact if the system is smooth one can imagine, and sometimes prove rigorously, see [RT98], that phase space contains *big nonchaotic islands* where the system looks very different from a Anosov system: it is important to understand the relevance of such regions for the statistical properties of (moderately) large systems because the chaotic hypothesis states that they should not be too relevant for nonequilibrium theory.

The *problem* that "remains" is whether the chaotic hypothesis has any power to tell us something about nonequilibrium statistical mechanics. This is the real, deep, question for anyone who is willing to consider the hypothesis. Of course *one consequence* is the ergodic hypothesis, hence the heat theorem, but this is manifestly *too little* even though it is a very important property for a theory with the ambition of being a *general* extension of the theory of equilibrium ensembles.

A chaotic motion, as discussed in §9.2, is recognized from the expansion and contraction that the evolution map $x \to Sx$ produces on line elements of phase space Σ (the space of the events).

However the notion of expansion and contraction depends on the metric that we use near x and Sx to measure lengths; hence it is clear that expansion and contraction at x *are not definable* in terms of x and of the action of S near x alone. The latter are, in fact, local notions, but it will make sense to say that a line element δ emerging from x (lying on Σ) "*expands*" *if it does so asymptotically*, *i.e.* if:

9.4.1 $$|S^n\delta|_{S^n x} > Ce^{\lambda n}|\delta|_x \qquad \text{for all } n > 0 \tag{9.4.1}$$

where $|\cdot|_x$ denotes the length of δ measured with the metric used at the point x of Σ and $C, \lambda > 0$ are suitable constants. Likewise one has to reason in the case of contraction.

The collection of the expanding line elements generates a plane T^u_x (tangent to Σ) called the *expanding plane* at x and the collection of the contracting line elements generates another plane T^s_x. This can be illustrated in a two-dimensional case by the first of Fig. 9.1.1 where two line elements emerging from a point $x \in \Sigma$ are directed along the expanding and the contracting directions.

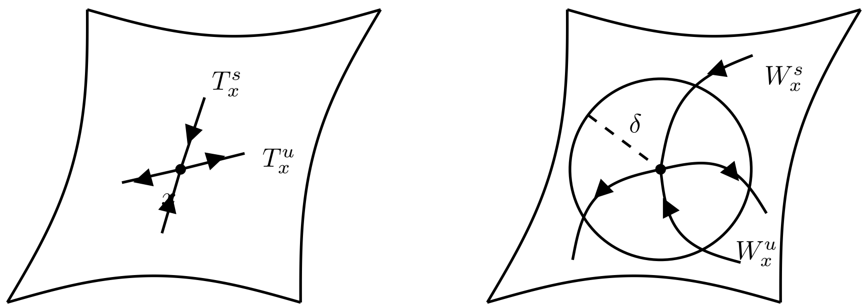


Fig. 9.1.1

Here the two tangent planes emerging out of x are drawn as two arrows: one has to imagine that the planes can be drawn at all points of Σ; in fact the family of planes T^s_x (or T^u_x) can be *integrated* and generates a family of smooth manifolds W^s_x (or W^u_x) that are tangent at each of their points y to the plane T^s_y (or T^u_y). The second drawing in Fig. 9.1.1 illustrates this property and shows the beginning of the two manifolds through x, *i.e.* a part of the manifolds that is enclosed in a sphere of radius δ, small compared to the curvature of the manifolds.

More generally we can hope that it will be possible to give a decomposition of the tangent plane T_x at a point x as a sum of linearly independent planes T^j_x:

9.4.2
$$T^1_x \oplus T^2_x \oplus \ldots \oplus T^p_x \tag{9.4.2}$$

of dimensions $n_1, \ldots, n_p$ and to define p *Lyapunov exponents* $\lambda_1, \ldots, \lambda_p$ such that:

9.4.3
$$\lim_{n\to+\infty} \frac{1}{n} \log |S^n\delta|_{S^n x} = \lambda_j \qquad \text{if } \delta \in T^j_x\,. \tag{9.4.3}$$

If such a decomposition and such exponents exist we say that the point x admits a *dynamical base* for the evolution S *"towards the future"*.

If ∂S^n_x denotes the matrix of the derivatives of S^n evaluated at x the the spaces $T^1_x, \ldots, T^p_x$ can be taken to be the eigenspaces of the matrix $\lim_{n\to+\infty} \big((\partial S^n_x)^*(\partial S^n_x)\big)^{1/2n}$, if this limit exists: *however* it should be clear that this is not the only way in which the tangent plane can be split so that (9.4.3) holds.

This is so because the $*$-operation (*i.e.* transposition) depends on the system of coordinates and on the metric: one sees this *by noting that* ∂S^n_x *maps the tangent space* T_x *into a different space*, namely $T_{S^n x}$. For instance if we have a decomposition satisfying (9.4.2) we get a new one by changing T^1_x into a new plane forming a different (positive) angle with respect to

the plane of the other vectors $T_x^2 \oplus \ldots \oplus T_x^p$; and one can likewise change $T_x^2, T_x^3, \ldots$.

But the sequence of subspaces T_x^p, $T_x^p \oplus T_x^{p-1}$, $T_x^p \oplus \ldots \oplus T_x^2$, $T_x \equiv T_x^p \oplus \ldots \oplus T_x^1$, is uniquely determined and metric independent (assuming the existence of a dynamical base towards the future).

Dynamical bases for the evolution towards the past are defined in the same way if S is invertible (we do not describe here the obvious adjustments needed in the definition, *e.g.* $n \to -\infty$). The same comments about lack of uniqueness can be made for such bases.

On the other hand in the case of invertible maps dynamical bases which are such both in the future and in the past may exist and, if they do exist, they are uniquely determined; we shall call them simply *dynamical bases* or *bilateral dynamical bases*.

It is remarkable that such bilateral bases exist under rather general conditions. For instance:

(a) if μ is a probability distribution which is *S-invariant* [7] then, apart from a set of points of zero μ-probability, every point admits a dynamical base, and

(b) if the distribution μ is *ergodic* in the sense that there are no nontrivial *measurable* functions that are constants of motion then the dimensions of the planes T_x^j, as well as the Lyapunov exponents λ_j, are x-independent with μ-probability 1.

Remarks:

(1) Properties in (a) and (b) are the content of *Oseledec's theorem*.

(2) Dynamical bases for the motion towards the future and those for the motion towards the past will be different, in general. However if S is invertible and μ is an *invariant distribution* the forward and backward dynamical bases can be chosen to coincide (apart for a set with zero μ-probability) and their exponents are *opposite*, see [Ru79a] p. 283.

(3) A *caveat* is that in (2) it is essential that μ be S-invariant. Therefore the above is *not* saying that all points but a set of zero volume will admit a dynamical base, because in general the volume measure μ_0 *is not invariant.* But if μ_0 admits a statistics μ in the sense of (9.3.1) then all points but a set of μ-measure 0 admit a dynamical base. In other words it is important to note that in general an invariant probability distribution μ is not expressible by means of a density function in phase space (one says that it is not "*absolutely continuous*" or "*nonsmooth*"), and therefore one cannot say that the latter statement (b) holds "apart from a set of points with zero volume", even when the volume distribution μ_0 admits a statistics μ.

---

[7] *i.e.* the probability $\mu(E)$ of E and that of $S^{-1}E$ are equal for all sets E's.

Considering a dynamical system the question of the existence of the stable and unstable manifolds or just of the local stable and unstable manifolds is of course rather delicate and one would like to have simple critieria guaranteeing their exixtence and a few basic properties. The real problem is with the *local* manifolds, $W_x^{s,\delta}, W_x^{u,\delta}$, in a sphere of radius δ around a point x. In fact once the local manifolds are defined one can define the *global expanding and contracting manifolds* even when S is not invertible, simply as

9.4.4
$$\begin{aligned} W_x^u =&\{\cup_{n=0}^\infty S^n W_{x_n}^{u,\delta} \quad \text{if } S^n x_n = x\} \\ W_x^s =&\{\text{set of points } y \text{ such that } S^n y \in W_{S^n x}^{s,\delta} \quad \text{for } n \text{ large enough}\} \end{aligned} \tag{9.4.4}$$

where the first union is meant, when S is not invertible, also as a union over the various possibilities of choosing the x_n's.

The existence of the local hyperbolic structure may seem a property difficult to check. This is ideed so; but in many cases the proof is greatly simplified because of the following sufficient conditions.

The existence of the local manifolds $W_x^{s,\delta}, W_x^{u,\delta}$ can be deduced from the existence of a continuous family of *cones* Γ_x^u, Γ_x^s lying in the tangent plane to every point x with the property that a displacement δ that points into the cone Γ_x^u will be transformed by the evolution map S inside a cone which is strictly less wide than the cone Γ_{Sx}^u and will have a length $|S\delta|$ which is strictly larger than that of δ, and a corresponding property holds for cone Γ^u, see Fig60.2 below.

Here "strictly" means, see Fig60.2, that the length will be larger by a factor $\overline{\lambda} > 1$ with respect to the initial length and $\overline{\lambda}$ will be independent of the point x; and, thinking that the cone is determined by its intersection with the unit sphere in T_x, "strictly less wide" will simply mean that the image $S\Gamma_x$ intersects the unit sphere of T_{Sx} in a set which has a distance $r > 0$ to the boundary of $\Gamma_{Sx} \cap T_{Sx}$, with r being x-independent.

The first of Fig60.2, illustrates the parts of the pair of cones Γ_x^u, Γ_x^s around x (shaded sectors) contained inside a small sphere around x. The evolution S maps them into the shaded sectors of the second figure, so that the expanding cone (marked by u) ends "well inside" the corresponding one around Sx (unshaded) while the contracting cone widens around the corresponding cone for Sx.

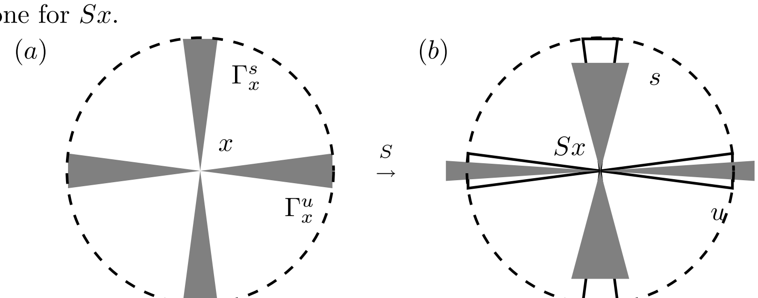


Fig. 9.4.2

S A similar property is required for the cones Γ_x^s, by using S^{-1} instead of S.

The existence of families of cones is sufficient for the system to be an Anosov system, but it is a much simpler condition, conceptually, and in fact it can even be easily satisfied in some cases, [Ru79], [Li95].

§**9.5. Symbolic Dynamics and Chaos**

The key property of mixing Anosov systems is that of admitting a "Markov's partition"; this is an important discovery that was realized in the classical works of Sinai, [Si68], [Si72], heralded by the independent discovery of a special case, [AW68]. We devote this section to the illustration of this important geometrical notion and to the consequent "symbolic dynamics", or "coarse grained" representation of motion.

The geometric properties of Anosov systems allow us to imagine a partition of phase space into *rectangular cells* $E_1, \ldots, E_{\mathcal{N}}$ such that, see Fig. 9.5.1:

(i) Each cell is defined by a "*center*" c and two "*axes*" Δ^u_c, Δ^s_c: it consists of the points $z(\xi,\eta)$ which have the form $z(\xi,\eta) = W^{u,\delta}_\eta \cap W^{s,\delta}_\xi \stackrel{def}{=} \xi\times\eta$ for some $\xi\in\Delta^u_c$ and $\eta\in\Delta^s_c$. The boundary ∂E of such a cell E will, therefore, consist of a *shrinking part* (or "stable part") $\partial^s E \stackrel{def}{=} \partial\Delta^u_c\times\Delta^s_c$ and of an *expanding* part (or "unstable part") $\partial^u E \stackrel{def}{=} \Delta^u_c\times\partial\Delta^s_c$.

The geometrical construction is illustrated *in the two-dimensional case* in Fig. 9.5.1. The circles are a neighborhood of $c\equiv x$ of size δ very small compared to the curvature of the manifolds (so that they look flat); Fig. 9.5.1a shows the axes; Fig. 9.5.1b shows the $\times$ operation and $W^{u,\delta}_\eta, W^{s,\delta}_\xi$ (the horizontal and vertical segments through η and ξ, respectively, have size δ); Fig. 9.5.1c shows the rectangle E with the axes (dotted lines) and with the four marked points being the boundaries $\partial\Delta^u$ and $\partial\Delta^s$. The picture refers to the two-dimensional case (which is substantially easier to draw and to conceive, see [Bo70]), and the stable and unstable manifolds are drawn as flat, *i.e.* the Δ's are very small compared to the curvature of the manifolds. Transversality of W^u_x, W^s_x is pictorially represented by drawing the surfaces at 90° angles:

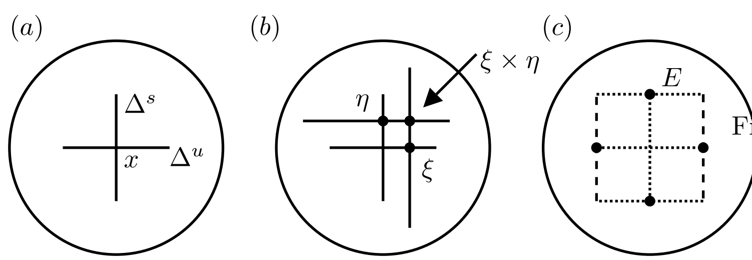


Fig. 9.5.1

(ii) Furthermore we require a *covariance* property of the various cells with respect to the action of the evolution map S: *i.e.* we demand that the map S transforms the shrinking parts of the boundary of E inside the union of the shrinking parts of the various cells of the partition, and S^{-1} also enjoys the corresponding property (with the collection of "expanding" sides of the

cells now containing the S^{-1} images of the "expanding" sides of each cell). The covariance property is illustrated by Fig. 61.2.

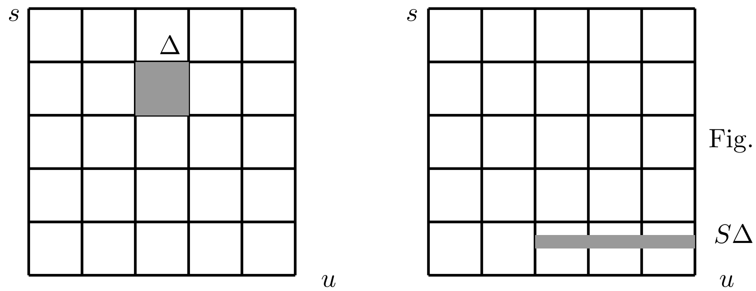


Fig. 9.5.2

A partition into such cells is called a *Markov partition*. Such partitions enjoy remarkable properties of covariance under the time evolution and are suitable for a description of the motion.

Note that here "shrinking" and "expanding" surface elements are represented in a "literal" sense. However since these are asymptotic notions (which only as such are metric independent) it may well be that under the action of S a "shrinking side" actually expands in the metric used, or an "expanding side" actually contracts: however under repeated applications of S such surface elements do eventually behave as the words, and drawings, we use suggest.

Before proceeding it seems useful to discuss briefly an example. It is the map of the two-dimensional torus $\mathcal{T} \equiv [0,2\pi]^2$ defined by $(x,y) \to (x+y, x+2y)$ (modulo 2π) or $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ (modulo 2π). This is the case in which Markov's partitions were first discovered, [AW68].

In this case we see easily that the expanding and contracting planes are simply the lines through a point (x,y) parallel to the eigenvectors of the matrix $M = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. The stable and unstable manifolds are, therefore, lines through (x,y) parallel to these eigenvectors and regarded as drawn on the manifold $\mathcal{T}$.

Thus they cover it densely because the slope of these lines is irrational (being $(1 \pm \sqrt{5})/2$). We see in this example also why it is (in general) necessary to distinguish between the local stable and unstable manifolds and the global ones (which are dense while the local manifolds are not).

It is easy, by using ruler and compass, to draw a Markov pavement for the above map: an example is given in Fig. 9.5.3.

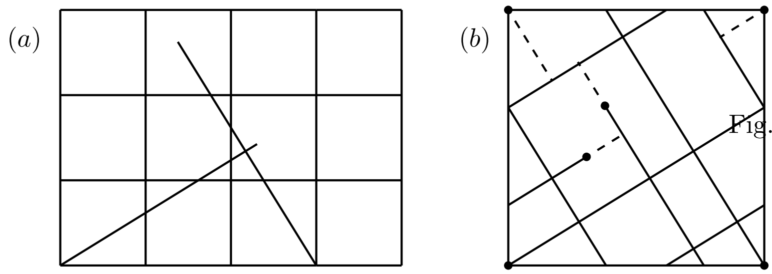


Fig. 9.5.3

This is a partition for the torus map just introduced: Fig. 9.5.3a represents many copies of the torus pavement in Fig. 9.5.3b (unraveled as a pavement of the plane). Each copy is a square with side 2π. The two lines are parallel to the eigenvectors of the above matrix M generating the map. The line with positive slope is parallel to the eigenvector with eigenvalue larger than 1, and the two lines are orthogonal because the matrix is symmetric.

In this simple case the expanding and contracting directions are trivially parallel to the two drawn lines. If we draw the lines of Fig. 9.5.3a as they really appear when wrapped back into the torus we get Fig. 9.5.3b where the endpoints of the lines are marked as big dots. The lines define several rectangles (both in the geometrical sense and in the above introduced *dynamical* sense, which explains the attribute "rectangular" used) with a few exceptions (three in the picture). They are due to the termination of the initial lines "in the middle of nowhere". The dashed lines continue the original lines (on both ends) until they meet a line, thus completing the missing rectangles. The result depends on the order of the continuation operations: an irrelevant ambiguity (as in any event Markov's partitions are by no means unique[8], when they exist).

Fig. 9.5.3b could have been obtained directly without any elongation of the lines had the initial lines been drawn of appropriate size. The construction shows how to find the appropriate size (using only "ruler and the compass", as required by every noble drawing, or by a Postscript program using only "integers and quadratic irrationals").

The union of the rectangle boundaries parallel to the line with negative (positive) slope is transformed into itself by the action of the map (inverse map). This is so because it is a connected piece of the stable manifold of the trivial fixed point that is the origin; it shows that the property of Fig. 9.5.2 is indeed satisfied. Hence Fig. 9.5.3b is a *simple example of a "Markovian pavement"* also called a "*Markov partition*" (and its discovery was at the beginning of the developments discussed here, [AW68], [Si68]).

In fact given a Markov partition we can generate a much finer partition simply by transforming it with the various iterates of the map and then intersecting the collection of pavements thus obtained, see footnote 8. The new partition of phase space is obviously *still Markovian but it can be made*

[8] For instance if $\mathcal{P} = \{E_i\}$ is a Markov partition then $\mathcal{P}' = \{E_i \cap SE_j\}$, $\mathcal{P}'' = \{S^{-1}E_i \cap E_j\}$ and $\tilde{\mathcal{P}} = \{S^{-1}E_i \cap E_j \cap SE_m\}$ (and so on) are Markov partitions.

as fine as we want.

The above definitions and related properties can be extended to systems which have an attracting set consisting of a smooth surface and the time evolution restricted to the surface becomes a mixing Anosov system. The extension can even be pushed to systems with attracting sets that are not smooth but that are still hyperbolic ("axiom A attractors", [Ru76]). We shall not deal here with this more general notion (for simplicity).

Anosov systems that are mixing are *topologically mixing* in the sense that given any pair of open sets U, V there is a n_V such that the iterates $S^n U$ have a nonempty intersection with V for all $n \geq n_V$. Hence imposing mixing is the simple way of imposing the condition that the system phase space cannot be "disconnected" into disjoint parts under the action of S or of any of its iterates (alternative to assuming existence of a dense orbit).

A general theorem ("*Smale's spectral theorem*") deals with cases in which one does not assume mixing: under the assumptions (a) and (b) in the definition of §9.4 and adding the extra assumption:

(c) Periodic motions are dense on phase space.

the phase space M can be decomposed into a union of $M_1 \cup M_2 \cup \ldots \cup M_n$ in each of which there is a dense orbit. If $n = 1$ the system is said to be *transitive*. Hence each system satisfying assumptions (a), (b), (c) can be regarded as a collection of finitely many transitive Anosov systems. Furthermore if a Anosov system is transitive then its phase space M can be decomposed as a union $M = M'_1 \cup M'_2 \cup \ldots \cup M'_m$ on each of which S^m acts as a mixing Anosov system.

In other words given an Anosov system with dense periodic points either topological mixing holds for the map S^n for some n, or phase space splits into a finite number of disjoint closed components (called "spectral elements" of the system) in each of which, for some large enough n, the map S^n is topologically mixing. The mixing assumption is not as strong as it may at first appear: if it does not hold it is because in some sense we have chosen the phase space inappropriately not noticing that motion was actually taking place on a smaller space.

Markov partitions set up a nice "*coarse graining*" permitting us to think of the dynamical system as a copy of something very familiar in statistical mechanics: namely the one-dimensional Ising model, or of one of its extensions considered in §5.10. The correspondence is via the *symbolic dynamics* associated with Markov partitions.

One defines for each point x a sequence $\underline{\sigma}(x)$ of digits each of which can take $\mathcal{N}$ values if $\mathcal{N}$ is the number of elements of a Markov partition $\mathcal{E}$ with elements so small that the image of every rectangle intersects all the other rectangles at most in a connected part (*i.e.* the size of the rectangles is so small that even when stretched by the *one time step* evolution map, it remains small compared to the curvature of the sides).

One defines a $\mathcal{N}\times\mathcal{N}$ *transition matrix* $T_{\alpha\beta}$ by setting $T_{\alpha\beta}=1$ if the interior of the rectangle E_α evolves under S into a set intersecting the interior of the rectangle E_β; we set $T_{\alpha\beta}=0$ otherwise.

The definitions readily imply a few properties. A transition matrix is said to be *transitive* if for any pair σ,σ' there is a power n such that $(T^n)_{\sigma,\sigma'}>0$; this means that there is a sequence which is compatible and which contains the symbols σ and, to the right of it, σ'. A transition matrix is said to be *mixing* if for all n large enough $(T^n)_{\sigma,\sigma'}>0$. A Markov partition for a transitive Anosov system has a transitive matrix which is also mixing if the system is mixing.

A sequence $\{\sigma_k\}$ is "*allowed*" or "*compatible*", see §5.10, if it is a sequence of symbols such that $T_{\sigma_k\sigma_{k+1}}\equiv 1$ for all integers $k\in(-\infty,\infty)$. A meditation on Fig. 9.5.3, *i.e.* on the *covariance* property, will convince the reader that if $\underline{\sigma}=\{\sigma_i\}$ is allowed there must be a point x such that $S^kx\in E_{\sigma_k}$ for all integers $k\in(-\infty,\infty)$; and this point must be unique by the hyperbolicity of the transformation (if there were two such points they would travel visiting always the same boxes of $\mathcal{E}$: which is impossible because if two points visit the same boxes during the time $[-T,T]$ then, from the definitions of hyperbolicity, their distance must be not bigger than $O(Ce^{-\lambda T})$, cf. (9.4.1)).

Therefore it follows that we can establish a correspondence between points and compatible sequences. There may be exceptionally *more* than one sequence representing the same point, but this can happen only if the point is either on a boundary of a rectangle of $\mathcal{E}$ or on that of one of its images under iterates of S. Hence the set of points that are represented by more than one sequence has zero volume and, therefore, it can be ignored for the purposes of our discussion (which in any event disregards sets of zero volume).

This means that we can map, or "*code*", phase space into a space of sequences: any function on phase space becomes a function of the sequences. The coding of points into sequences of digits is very similar to the familiar coding of its coordinates into decimal sequences (which is *also* well defined apart from a zero volume (dense) set of exceptional points, namely the points whose coordinates are numbers ending with an infinite string of 0's or of 9's). And *it is harder but much better* in spite of the fact that the decimal representation is the "usual" representation of points in phase space, both in theoretical applications and in numerical experiments.

It is better because it is *adapted to the dynamics* and turns the most chaotic dynamics into a "*standard*" one (still chaotic), namely the shift on a space of sequences of symbols subject to a nearest neighbor constraint (that $T_{\sigma_i,\sigma_{i+1}}=1$), also called a *hard core*, see §5.10. The latter dynamical systems are often called "subshifts of finite type", or "one-dimensional spin (or particle) chains" for obvious reasons.

A function $F(x)$ which is mildly regular on phase space, *e.g.* Hölder continuous with exponent a, becomes a function $F(x(\underline{\sigma}))$ of $\underline{\sigma}$ *which has a weak dependence* on the digits σ_k of $\underline{\sigma}$ with large $|k|$: *i.e.* if $\underline{\sigma}$ and $\underline{\sigma}'$ agree on the digits between $-k$ and k then the distance between $x(\underline{\sigma})$ and $x(\underline{\sigma}')$ is

$\le O(Ce^{-\lambda k})$ so that $F(\underline{\sigma}) - F(\underline{\sigma}')$ is bounded by a constant times $e^{-\lambda a k}$.

In particular the expansion rate $\Lambda_u(x)$ of the phase space volume over one time step of the map S (*i.e.* the determinant of the Jacobian ∂S of the map S, also called the *contraction rate of phase space* because $\Lambda_u(x)$ is a "real" contraction or a "real" expansion depending on whether it is < 1 or > 1) will satisfy for some $\overline{C}$, $\overline{a}$:

9.5.1 $$|\Lambda_u(x(\underline{\sigma})) - \Lambda_u(x(\underline{\sigma}'))| < \overline{C} e^{-k\,\overline{a}} \tag{9.5.1}$$

if the digits of $\underline{\sigma}$ and $\underline{\sigma}'$ with labels between $-k$ and k agree. This is because hyperbolicity implies that the stable and unstable tangent planes at x are very smooth and vary in a Hölder continuous fashion with the point x (as mentioned in §9.3, this is Anosov's theorem), so that also the functions $\Lambda_u(x), \Lambda_s(x)$ are Hölder continuous.

One can interpret (9.5.1) as saying that the function $\Lambda_u(x(\underline{\sigma})) \stackrel{def}{=} e^{\lambda(\underline{\sigma})}$ has *short range* as a function of the symbolic sequence $\underline{\sigma}$ and the following remarks are worth the effort necessary for understanding their formulation, admittedly hard at first sight (yielding to a reassuring sense of triviality after some thought):

(1) The set of points which are symbolically represented by sequences that agree between $-\frac{1}{2}T$ and $\frac{1}{2}T$ is just the set of points in the very small rectangle of $\mathcal{E}_{-\frac{1}{2}T,\frac{1}{2}T}$ consisting of the intersections $\cap_{-\frac{1}{2}T}^{\frac{1}{2}T} S^{-j} E_{\sigma_j}$ (with sides bounded proportionally to $e^{-\lambda\frac{1}{2}T}$), see footnote 8 above. Hence a sum over the elements $E \in \mathcal{E}_{-\frac{1}{2}T,\frac{1}{2}T}$ can be written as a sum over the sequences $(\sigma_{-k}, \ldots, \sigma_k)$ (which are compatible). The reader will be greatly helped by attempting to draw a representation of a set in $\cap_{-\frac{1}{2}T}^{\frac{1}{2}T} S^{-j} E_{\sigma_j}$ in the manner of Fig. 9.5.2 above).

(2) A point $x \in E = \cap_{-\frac{1}{2}T}^{\frac{1}{2}T} S^{-j} E_{\sigma_j}$ is determined by a compatible bi-infinite sequence $\underline{\sigma}$ which *continues* $\sigma_{-\frac{1}{2}T}, \ldots, \sigma_{\frac{1}{2}T}$ on either side to an infinite (compatible) sequence $\ldots, \sigma_{-\frac{1}{2}T-1}, \sigma_{-\frac{1}{2}T}, \ldots, \sigma_{\frac{1}{2}T}, \sigma_{\frac{1}{2}T+1}, \ldots$.

(3) Since there are only finitely many values for each symbol we can define for each symbol σ compatible sequences $\underline{\sigma}^R(\sigma) = (\sigma'_1, \sigma'_2, \ldots)$ and $\underline{\sigma}^L(\sigma) = (\ldots, \sigma'_{-2}, \sigma'_{-1})$ infinite to the right and to the left, respectively, and such that $T_{\sigma,\sigma'_1} = 1$ and $T_{\sigma'_{-1},\sigma} = 1$, *i.e.* respectively right and left compatible with σ: this can be done in many ways. We shall call a pair of functions $\sigma \to (\underline{\sigma}^R(\sigma), \underline{\sigma}^L(\sigma))$ a *boundary condition*.

If we are given a boundary condition $(\underline{\sigma}^R(\sigma), \underline{\sigma}^L(\sigma))$ we can associate with each element $E \in \cap_{-\frac{1}{2}T}^{\frac{1}{2}T} S^{-j} E_{\sigma_j}$ a point $c(E)$ whose (bi-infinite) symbolic sequence is $\underline{\sigma}^L(\sigma_{-\frac{1}{2}T}), \sigma_{-\frac{1}{2}T}, \ldots, \sigma_{\frac{1}{2}T}, \underline{\sigma}^R(\sigma_{\frac{1}{2}T})$. We call $c(E)$ the "*center*" of E with respect to the boundary condition $\sigma \to (\underline{\sigma}^R(\sigma), \underline{\sigma}^L(\sigma))$.

The bi-infinite continuation, which we can naturally call "*Markovian*", of the

finite string $\sigma_{-\frac{1}{2}T},\ldots,\sigma_{\frac{1}{2}}T$ to an infinite compatible sequence is possible because of the topological transitivity property supposed for the system (consequence of the mixing).[9]

(4) We can interpret a continuation of the sequence $\sigma_{-\frac{1}{2}T},\ldots,\sigma_{\frac{1}{2}T}$ as the assignment of a boundary condition to a spin configuration in the one-dimensional box $[-\frac{1}{2}T,\frac{1}{2}T]$ in the sense of the discussion in §5.10, following (5.10.16).

(5) If x is a point in $E=\cap_{j=-\frac{T}{2}}^{\frac{T}{2}} S^{-j}E_{\sigma_j}$ the product of expansion factors $\prod_{j=-\frac{T}{2}}^{\frac{T}{2}}\Lambda_u(S^{-j}x)^{-1}$ can be written

9.5.2
$$\prod_{j=-\frac{T}{2}}^{\frac{T}{2}}\Lambda_u(S^{-j}x)^{-1}=e^{\sum_{k=-\frac{T}{2}}^{\frac{T}{2}}\lambda(\vartheta^k\underline{\sigma})}\tag{9.5.2}$$

with $\lambda_a(\underline{\sigma})=\log\Lambda_a(x(\underline{\sigma}))$, $a=u,s$.

(6) One says that the mixing Anosov system (M,S) admits a symmetry if there is an isometry of phase space with $I^2=1$ which is either commuting or anticommuting with S: *i.e.* $IS=SI$ or $IS=S^{-1}I$, respectively. In the second case the symmetry is called a *time reversal* symmetry: of course the same system may admit several time reversal symmetries. A Markov pavement for (M,S) is said to be "I-symmetric" if $E\in\mathcal{P}$ implies that $IE\in\mathcal{P}$; in this case, again by the transitivity of the compatibility matrix, we can define *time reversible boundary conditions* $\sigma\to(\underline{\sigma}^R(\sigma),\underline{\sigma}^L(\sigma))$ in such a way that the centers $c(E)$ of the sets $E=\cap_{j=-\frac{T}{2}}^{\frac{T}{2}}S^{-j}E_{\sigma_j}$ satisfy the covariance property: $Ic(E)=c(IE)$.
In the commuting case this is so because we can choose the continuation of σ_k to the right and that of $i\sigma_k$ to the right to be "consistent" *i.e.* if $\sigma_{k+1},\sigma_{k+2},\ldots$ continues σ_k then $i\sigma_{k+1},i\sigma_{k+2},\ldots$ continues $i\sigma_k$. In the anticommuting case the continuation of σ to the right has to be chosen consistent in the same sense with that of $i\sigma$ to the left. Such consistent choices are Markovian in the above sense.

(7) another interesting consequence of mixing is that given a sequence $\sigma_{-s},\ldots,\sigma_s$ and supposing that p is such that $T^p_{\sigma,\sigma'}>0$ for all σ,σ' (which

[9] Given a symbol $\overline{\sigma}$, let $\overline{n}$ be such that $(T^{\overline{n}})_{\overline{\sigma},\overline{\sigma}}>0$. Let n_σ be such that $(T^{n_\sigma})_{\sigma,\overline{\sigma}}>0$, this means that there is a sequence $\overline{\sigma},\ \sigma'_2,\ \ldots,\sigma'_{\overline{n}-1},\overline{\sigma}$ which is compatible, and also a sequence $\sigma\tilde\sigma_2\ldots\tilde\sigma_{n_\sigma}\,\overline{\sigma}$ which is compatible: therefore $\underline{\sigma}^R(\sigma)\stackrel{def}{=}\tilde\sigma_2,\tilde\sigma_{n_\sigma},\,\overline{\sigma},\sigma'_2,\ldots,\sigma'_{\overline{n}-1},\ldots$ where the last dots indicate indefinite repetition of the string $\overline{\sigma},\sigma'_2,\ldots,\sigma'_{\overline{n}-1}$ is a infinite string "continuing" the symbol σ to the right into a compatible sequence. Likewise one builds a $\underline{\sigma}^L(\sigma)$ continuing th symbol σ to the left into a compatible sequence. The name "Markovian" is due to the property that the sequences $\underline{\sigma}^L(\sigma),\underline{\sigma}^R(\sigma)$ share: namely they depend solely on σ.

is an analytic form of the mixing property) we can form a compatible sequence $\sigma_{-s}, \ldots, \sigma_s, \sigma_{s+1}, \ldots, \sigma_{s+p} \equiv \sigma_{-s}$: therefore we can form an infinite sequence in which the latter sequence is repeated indefinitely into an infinite *and compatible* sequence. The point x that corresponds to this sequence is, necessarily, a periodic point (with period $s+p$). A consequence is the density of periodic points in phase space: every mixing Anosov system admits a dense set of periodic orbits.

Because of all the above properties Anosov systems (and the related Axiom A attractors) are the *paradigm* of chaotically behaving systems. The basic idea, developed in the early 1970s by D. Ruelle, is that such systems are not just curiosities but rather they are in some sense the "rule" when dealing with real dynamics.

§9.6. Statistics of Chaotic Attractors. SRB Distributions

Perhaps the most important property of mixing Anosov systems or systems with Axiom A attractors is that they admit stationary states, *i.e.* the limits in (9.3.1) exist and the "statistics" μ of *almost all* data x (*i.e.* all data outside a zero volume set) exists. Furthermore the probability distribution μ, called the *SRB-distribution*, describing them can be *characterized* quite explicitly.

Note that the "almost all" *is an essential feature of the definition, in fact,* see §9.4*, Anosov (or Axiom A) systems will have a dense set of periodic points covering phase space (or the attracting set): any such point x, at least, will of course be an exceptional point as far as the value of the limit in (9.3.1) is concerned.*

The SRB distribution can be given an expression in terms of the kinematical properties discussed in the previous section. This is an expression that can play a role similar to that played by the Boltzmann-Gibbs expression for the equilibrium distributions. It is an expression that clearly cannot be computed in any nontrivial case, much like the integrals that express equilibrium properties in terms of integrals with respect to the canonical distribution.

However, like the integrals with respect to the canonical distribution, it can be useful to derive relations that must hold between various averages. Therefore formal expressibility of the SRB distribution seems to be a *very important property* for nonequilibrium theory.

The formula can be rather easily justified at an informal level, however even this requires good will on the part of the reader, to the extent that he will develop it only if convinced of its utility. Therefore we relegate to Appendix 9.A2 below the "informal" analysis and we confine ourselves to giving here the SRB distribution expression in a form sufficient for the discussion of a few applications.

One needs to define the SRB average of a generic observable $F(x)$ on phase

space: $\langle F\rangle = \int F(x)\,\mu(dx)$. For this purpose we divide phase space into cells at least so small that $F(x)$ is constant in each of them.

The division into cells will be conveniently made by using a Markov partition $\mathcal{E}$ (it will not matter which one); of course the elements $E \in \mathcal{E}$ will not, in general, be so small that F is constant in each of them. However we can refine $\mathcal{E}$ simply by considering the partition $\mathcal{E}_{-T,T} = \cap_{j=-T}^{T} S^{-j}\mathcal{E}$; as remarked in §9.4 this is a much finer partition (and the size of the cells is of the order $O(e^{-\lambda T})$ if λ is defined in (9.4.1)).

Each element of $\mathcal{E}$ has the form $\cap_{j=-T}^{T} S^{-j} E_{\sigma_j}$ where $\sigma_{-T},\ldots,\sigma_T$ is a compatible sequence of symbols.

Given a boundary condition $\underline{\sigma}^L, \underline{\sigma}^R$ in the sense of §9.5, which one again does not matter, we can define a *compatible* bi-infinite sequence:

9.6.1
$$\underline{\sigma}_E \stackrel{def}{=} (\underline{\sigma}^L(\sigma_{-T}), \sigma_{-T},\ldots,\sigma_T, \underline{\sigma}^R(\sigma_T)) \tag{9.6.1}$$

for each $E \in \mathcal{E}_T$.

If $c(E)$ is the point of E whose symbolic representation relative to the partition $\mathcal{E}$ is $\underline{\sigma}_E$ we can define the *expansion rate* $\Lambda_{u,2T}(E)$ of the map S^{2T} regarded as a map between $S^{-T}c(E)$ and $S^T c(E)$ as, cf. (9.5.2),

9.6.2
$$\Lambda_{u,2T}(E) \equiv \Lambda_{u,2T}(c(E)) \stackrel{def}{=} \prod_{k=-T}^{T-1} \Lambda_u(S^{-k}c(E)) = \prod_{k=-T}^{T-1} e^{\lambda_u(S^{-k}c(E))}\,. \tag{9.6.2}$$

Then the SRB distribution μ can be written as

9.6.3
$$\int F(y)\mu(dy) = \lim_{\substack{T\to\infty, n(T)\to\infty\\ n(T)\le T}} \frac{\sum_{E\in\mathcal{E}_{-T,T}} \Lambda^{-1}_{u,2n(T)}(c(E))F(c(E))}{\sum_{E\in\mathcal{E}_{-T,T}} \Lambda^{-1}_{u,2n(T)}(c(E))} \tag{9.6.3}$$

where $n(T) \le T$ is any sequence tending to ∞ as $T\to\infty$.

A particularly convenient choice will be $n(T) = T$ so that, with $c \equiv c(E)$,

9.6.4
$$\int F(y)\mu(dy) = \lim_{T\to\infty} \frac{\sum_{E\in\mathcal{E}_{-T,T}} \Lambda^{-1}_{u,2T}(c)F(c)}{\sum_{E\in\mathcal{E}_{-T,T}} \Lambda^{-1}_{u,2T}(c)} \tag{9.6.4}$$

Remarks:

(1) The weight in (9.6.4) can be written, in terms of the function $\lambda(\underline{\sigma})$ defined in §9.5, see (9.5.2), and of the *shift* operation ϑ on the infinite sequences, as

9.6.5
$$\Lambda^{-1}_{u,n(T)}(E) = e^{-\sum_{j=-n(T)}^{n(T)} \lambda(\vartheta^j\underline{\sigma})} \tag{9.6.5}$$

where $\underline{\sigma} = \underline{\sigma}_E$ is the sequence that is obtained by continuing $\sigma_{-T},\ldots,\sigma_T$ to an infinite sequence as prescribed by the chosen boundary condition, see (9.6.1). This is a slightly different rewriting of (9.6.2). It shows that

the SRB distribution can be interpreted as a probability distribution on the space of bi-infinite compatible sequences $\underline{\sigma}$. As such it is a *Gibbs distribution* for a one-dimensional lattice spin model *with short-range interaction* and nearest neighbor hard core (due to the compatibility restriction that limits the allowed configurations $\underline{\sigma}$). In fact (9.5.1) says that $\sum_{j=-n(T)}^{n(T)} \lambda(\vartheta^j \underline{\sigma})$ can be interpreted as the energy of a spin configuration under a potential that is exponentially decreasing at ∞, see §5.10, (5.10.14), (5.10.18).

In §5.8 we have seen that one-dimensional short-range systems are quite trivial from the statistical mechanics viewpoint and the theory of chaotic motions inherits quite a few results from the theory of such lattice systems. Such results are often quite nontrivial and surprising when seen as properties of chaotic motions. For instance uniqueness of SRB distributions corresponds to the absence of phase transitions in one-dimensional lattice systems. The thermodynamic limit corresponds to the limit as $T \to \infty$. Exponential decay of time correlations in the SRB distributions corresponds to the exponential decay of correlations in one-dimensional short-range lattice systems. Large deviation theorems correspond to the analyticity of the thermodynamic functions and so on.

For the above reasons the theory of Anosov (and Axiom A) systems has been called *thermodynamic formalism*, [Ru78b].

(2) If $n(T)/T \xrightarrow[T\to\infty]{} 0$ then by (9.6.5) and (9.5.1) we see that the weight $\Lambda^{-1}_{u,2n(T)}(E)$ given to the cell E of $\mathcal{E}_T$ *does not change appreciably* as the boundary condition is changed.[10] This is a kind of "mean value theorem", for the SRB distribution.

(3) But if $n(T) = T$ the variation of $\Lambda^{-1}_{u,2T}(E)$ within E *is appreciable because, clearly, the sum in (9.5.2) will undergo variations of order* $O(1)$, when the boundary condition is changed.

(4) Hence (9.6.4) is a deeper property than (9.6.3) with $n(T)/T \to 0$. It is proved easily in the thermodynamic formalism because it reduces to the statement that one-dimensional lattice gases with short-range interactions show no phase transitions, therefore the boundary condition dependence of the averages of local observables disappears in the thermodynamic limit $T \to \infty$), see §5.8.

(5) If the attractor is invariant under the action of a time reversal symmetry I, see §9.5, we can, and shall, suppose that the Markov partition is I-reversible: if $E \in \mathcal{E}$ then $IE \in \mathcal{E}$.[11] Furthermore the centers c in (9.6.2) can be chosen so that if c is the center of E then ic is that of IE, see §9.5, comment (6). This means that we have to choose a reversible boundary

---

[10] Note that at fixed E and as the boundary condition is varied the center point $c(E)$ varies (densely) inside E.

[11] If not one could use the finer partition obtained by intersecting $\mathcal{E}$ and $I\mathcal{E}$ so that the new partition will be time reversible.

condition to select the centers.

An expression of the SRB distribution based on a reversible partition $\mathcal{E}$ and on a reversible boundary condition will be called a *reversible representation of a SRB distribution.*.

In Appendix 9.A2 below we give details of an informal derivation of (9.6.3), but the above properties will suffice for deducing many interesting consequences of the chaotic hypothesis, in the form of general properties of Anosov systems.

It is interesting and important to discuss the connection between the Boltzmanian representation of motion as a cyclic permutation of the phase space cells and the new symbolic representation of motion in Anosov systems. In fact the chaotic hypothesis is supposed to hold also in the equilibrium cases and therefore one has *two different representations* of motions for the same system.

According to views already repeatedly expressed by Boltzmann, [Bo74], this "dualism" is not *a priori* impossible although in most cases in which he envisaged this possibility it seems that he did not really believe in it. The above seems to be a very fine and nontrivial instance in which a *dual representation* is possible. This is discussed in Appendix 9.A1 below, see also [Ga95a].

§**9.7. Entropy Generation. Time Reversibility and Fluctuation Theorem. Experimental Tests of the Chaotic Hypothesis**

The connection between the general kinematical analysis of chaotic motions and applications can be established if one accepts that the motions of a many-particle system are so "chaotic" that one can regard the system as a mixing Anosov system in the sense of §9.4.

One of the key notions in equilibrium statistical mechanics is that of *entropy*; its extension to nonequilibrium is surprisingly difficult, assuming that it really can be extended. In fact we expect that, in a system that reaches under forcing a stationary state, entropy is produced at a constant rate so that there is no way of defining an entropy value for the system, except perhaps by saying that its entropy is $-\infty$.

Although one should keep in mind that there is no universally accepted notion of entropy in systems out of equilibrium, even when in a stationary state, we shall take the attitude that in a stationary state only the entropy *creation rate* is defined: the system entropy decreases indefinitely, but at a constant rate.[12] Note that we say "decreases" and not "increases" because in a nonequilibrium situation nonconservative forces work upon the system and, since the system is supposed to be in a stationary sate, such work must be ceded to the exterior in the form of heat at constant temperature. So

[12] Defining "entropy" and "entropy production" should be considered an open problem.

entropy (of the system) decreases and entropy of the surroundings increases (as expected).

The natural definition of entropy creation rate is, following Gibbs and Boltzmann, that of the time derivative of the entropy of a state of the system that evolves towards stationarity, [An82], [Ru97c].

We look, temporarily, again at the evolution of our systems in continuous time: so that we imagine that it corresponds to a differential equation $\dot x = f(x)$. A "state" which at time $t=0$ is described by a distribution μ_0 with density $\rho_0(x)$ with respect to the volume element dx on phase space becomes at time t a distribution $\mu_t(dx) = \rho_t(x)dx$ with density $\rho_t(x) = \rho_o(S_{-t}x)\frac{\partial S_{-t}x}{\partial x}$, where $J_t(x) \stackrel{def}{=} \frac{\partial S_{-t}x}{\partial x}$ is the Jacobian determinant at x of the map $x \to S_{-t}x$. Defining the "entropy" of μ_t as

9.7.1
$$\begin{aligned}\mathcal{E}(t) &\stackrel{def}{=} -\int \rho_t(x)\log\rho_t(x)\,dx \equiv\\ &= -\int \rho_0(S_{-t}x)\,\frac{\partial S_{-t}x}{\partial x}\,\log\Big(\rho_0(S_{-t}x)\frac{\partial S_{-t}x}{\partial x}\Big)\,dx\end{aligned} \tag{9.7.1}$$

we deduce that

9.7.2
$$\begin{aligned}\mathcal{E}(t) = -\int \rho_0(S_{-t}x)\frac{\partial S_{-t}x}{\partial x}\log\rho_0(S_{-t}x)\,dx-\\ -\int \rho_0(S_{-t}x)\frac{\partial S_{-t}x}{\partial x}\log\frac{\partial S_{-t}x}{\partial x}\,dx\end{aligned} \tag{9.7.2}$$

and the first term on the righ-hand side does not contribute to $\dot{\mathcal{E}}$ because it equals the constant $-\int\rho_0(y)\log\rho_0(y)dy$ (just set $S_{-t}x = y$); therefore $\dot{\mathcal{E}}$ equals the t derivative of the second term, which can be transformed by setting $y = S_{-t}x$ into

9.7.3
$$-\int dy\,\rho_0(y)\,\log\frac{\partial S_{-t}(S_t y)}{\partial S_t y} = \int dy\,\rho_0(y)\,\log\frac{\partial S_t y}{\partial y} \tag{9.7.3}$$

having used the identity

9.7.4
$$\frac{\partial S_{-t}(S_t y)}{\partial S_t y}\frac{\partial S_t y}{\partial y} \equiv 1. \tag{9.7.4}$$

We now make use of the other identity

9.7.5
$$\frac{d}{dt}\frac{\partial S_t y}{\partial y} = -\frac{\partial S_t y}{\partial y}\,\sigma(S_t y) \tag{9.7.5}$$

where $\sigma(x)$ is the *divergence* of $-f(x)$ (writing the equations of motion as $\dot x = f(x)$). It follows that the *rate of entropy creation* is, see [An82], $\dot{\mathcal{E}}$:

9.7.6
$$\begin{aligned}\dot{\mathcal{E}} &= -\int \rho_0(y)\,\sigma(S_t y)\,dy =\\ &= -\int \rho_0(S_{-t}z)\frac{\partial S_{-t}z}{\partial z}\,\sigma(z)\,dz =\\ &= -\int \mu_t(dz)\sigma(z) \xrightarrow[t\to+\infty]{} -\int \mu(dz)\,\sigma(z) \stackrel{def}{=}\\ &\stackrel{def}{=} -\sigma_+\end{aligned} \tag{9.7.6}$$

if μ_t suitably converges to μ.

Hence we see that if $\mu_t \xrightarrow[t\to+\infty]{} \mu$ the asymptotic average rate of entropy variation is $\sigma_+ \stackrel{def}{=} \langle\sigma\rangle_\mu$. Since the macroscopic state of the system does not change as $t \to \infty$ we must interpret σ_+ as the average entropy increase of the thermostat that absorbs the heat created in the system by the forcing.

Naturally we must expect $\sigma_+ > 0$. It is reassuring that this a theorem, [Ru96a], that $\sigma_+ \ge 0$ for systems with an Axiom A attractor, hence for systems satisfying the stronger chaotic hypothesis. This is not very surprising because, by assumption, our system evolves in a bounded region (*i.e.* phase space is bounded) so that $\sigma_+ < 0$ would mean that the volume *expands* indefinitely, which is impossible.

Because of the above considerations we shall call σ the *entropy generation rate* and we suppose that it has the property $\sigma(x) = 0$ if the system is not subject to forcing, so that at zero forcing the evolution is volume preserving (a property usually true because the nonforced system is, as a rule, Hamiltonian).

Coming back to our previous point of view, with time evolution described by a map S on a phase space of "timing events" the entropy creation rate will be, in this case, identified with the phase space contraction between one timing event and the next:

9.7.7
$$\sigma(x) = -\log|\frac{\partial S(x)}{\partial x}|. \tag{9.7.7}$$

Our analysis concerns idealized systems of the above type that are also mixing *Anosov maps* in the sense of §9.4.

We now attempt to deduce *other* consequences of the chaotic hypothesis, possibly new (and in any event beyond the existence of the stationary distribution μ and the nonnegativity of σ_+) *and measurable* in at least some simple cases. The simplest cases to study are systems whose dynamics is reversible not only in the nonforced case, but under forcing as well.

Examples of thermostatting mechanisms that generate reversible motions are provided under rather general circumstances by forces acting on otherwise Hamiltonian systems and realizing an anholonomic constraint according to the principle of *least constraint* of Gauss, also called *minimal constraint principle*see Appendix 9.A4. In the following we shall provide some simple concrete examples, but it is important to note that the theory is far more general than the few examples that we shall discuss.

We consider, therefore, a general reversible mechanical system governed by a smooth equation:

9.7.8
$$\dot{x} = f(x, \underline{G}) \tag{9.7.8}$$

depending on several parameters $\underline{G} = (G_1, \ldots, G_n)$ measuring the strength of the forces acting on the system and causing the evolution $x \to S_t x$ of the phase space point x representing the system state in the *phase space* $\mathcal{F}$ which can be, quite generally, a smooth manifold.

We suppose that the system is "thermostated" so that motions take place on bounded smooth invariant surfaces $H(x;\,\underline{G}\,) = E$, which are level surfaces of some "level function" H. Hence we shall identify, to simplify the notation, the phase space $\mathcal{F}$ with this level surface which we shall sometimes call, somewhat inappropriately, the "energy surface".

We suppose also that the flow S_t generated by (9.7.8) is *reversible*, *i.e.* there is a volume preserving smooth map I, "time reversal", of phase space such that $I^2 = 1$ and "anticommuting with time":

9.7.9
$$S_t I = I S_{-t} \tag{9.7.9}$$

i.e. $f(Ix,\,\underline{G}\,) = -(\partial_x I)^{-1}(x)\cdot f(x,\,\underline{G}\,)$.[13]

We shall further restrict our attention to mixing Anosov systems that are *reversible*, in the above sense, for all values of the forcing parameters $\underline{G}$ of interest and *dissipative* at $\underline{G}\neq\underline{0}$. This means that the systems we consider are such that, see (9.7.6):

9.7.10
$$\sigma = \langle\sigma\rangle_\mu > 0 \qquad \text{for } \underline{G}\neq\underline{0} \tag{9.7.10}$$

Under the above assumptions one can define, for $\langle\sigma\rangle_\mu > 0$, the "dimensionless average entropy creation rate" p by setting:

9.7.11
$$p = \frac{1}{\langle\sigma\rangle_\mu}\frac{1}{\tau}\int_{-\tau/2}^{\tau/2}\sigma(S_t;\,\underline{G}\,)dt \tag{9.7.11}$$

Then the probability distribution of the variable p with respect to the SRB distribution μ can be written for large τ as $\pi_\tau(p)dp = \mathit{const}\, e^{-\tau\zeta_\tau(p)}dp$, see [Si77], and the function $\zeta(p) = \lim_{\tau\to\infty}\zeta_\tau(p)$ satisfies, if $\sigma_+ \equiv \langle\sigma\rangle_\mu > 0$ and $|p| < p^*$ for a suitable $p^*\geq 1$, the property:

9.7.12
$$\zeta(-p) = \zeta(p) + p\sigma_+, \qquad |p| < p^* \tag{9.7.12}$$

which is called the *fluctuation theorem*, and is part of a class of theorems proved in [GC95], see also [Ga95a], for discrete time systems, and in [Ge98], for continuous time systems. This theorem can be considerably extended, as discussed in [G96b], [Ga98b] and the extension can be shown to imply, in the limit $\underline{G}\to\underline{0}$ (when also $\sigma_+\to 0$) relations that can be identified in various cases with Green-Kubo's formulae and Onsager's reciprocal relations, see also [GR97], [Ga98d] and §9.9 below.

Similar theorems can be proved for suitable nonstationary probability distributions and, in fact, preceded the above, [ES94], or for nondeterministic evolutions, [Ku97], [LS98]. In the closest cases the relations between the

---

[13] For instance if $-\sigma(x;\,\underline{G}\,)$ is the rate of change of the volume element of $\mathcal{F}$ near x and if $\mathcal{F}$ is a Euclidean space then $\sigma(x) = -\sum_\alpha \partial_\alpha f_\alpha(x;\,\underline{G}\,)$. If $x = (p,q)$, and $Ix = (-p,q)$ and f is an Hamiltonian part plus a p-dependent term due to the "thermostat forces" then $\sigma(Ix,\underline{G}) = -\sigma(x,\underline{G})$.

latter theorems and the above is sometimes "*only*" an interchange of limits: it is precisely in the analysis of this interchange that the chaotic hypothesis plays a major role, see [CG99].

The interest of (9.7.12) is its *universal nature*, *i.e.* its (reversible) system independence, and the fact that it contains *no free parameter*.

The connection with applications of the above results is made via the assumption that concrete chaotic dynamical systems can be considered, "for the purpose of studying macroscopic properties", as mixing Anosov flows.

The fluctuation theorem proof is quite simple and it is based on the expression (9.6.4): it will be discussed in the forthcoming sections.

As a concluding comment we note that the probability distribution of p can be regarded as the probability distribution of the sum of the "local Lyapunov exponents"; if one defines it as the sum of the eigenvalues $\lambda_j(x)$ of the matrix $\partial S^\tau(S^{-\frac{\tau}{2}}x)$: $\tau p\langle\sigma\rangle_+ = -\sum_j \lambda_j(x)$.

If a system is an Anosov system then it has been proved that the probability distribution of the sum $p\sigma_+\tau$ of the local Lyapunov exponents has a density of the form $e^{\zeta(p)\tau}$, [Si77]. One says that the distribution is *multifractal* if $\zeta(p)$ is not linear. This means that the sum of the local Lyapunov exponents has wide fluctuations around its average (given by $\sim \tau\sigma_+$). The fluctuation theorem says that *the odd part of the multifractal distribution is always linear, in reversible systems*, so that multifractality of the volume contraction rate is, in such systems, related to the even part of its distribution.

Is the above (9.7.12) an observable relation? In fact it was observed in a numerical experiment with 56 particles modeling a (reversible) gas in a shear flow, [ECM93], and the attempt at theoretical prediction of the observed results led to the chaotic hypothesis and to the derivation discussed in §9.9.

It has then be observed in a sequence of experiments with 2 and 10 hard core particles, [BGG97], moving among fixed obstacles in a periodic box and subject to a constant field and thermostatted with a force necessary to maintain a constant total kinetic energy in spite of the action of the field. The force is selected among the several possible force laws as the one satisfying Gauss' principle of minimal constraint (so that the resulting equations are reversible, see Appendix 9.A4).

One can also consider systems in which the forcing has a "thermal nature" like systems enclosed in boxes whose walls are kept at constant temperature (depending however on which side of the walls one considers). Also such systems can be modeled with equations of motion which can be reversible, for instance see [Ga96b]. A very interesting numerical experiment has been performed on a chain of oscillators (in number up to 10^4) interacting with inelastic forces and with the oscillators at the extremes forced to have a "given temperature" by acting on them with suitable forces, [LLP97].

An experiment of a completely different kind, on a sample of water in convective chaotic motion (not too strongly chaotic, however) has been performed recently. Its interpretation in terms of the fluctuation theorem (or

rather of its extensions discussed in [Ga96c], [Ga97a], because a fluid is strongly dissipative and one cannot expect that the attractor is dense in phase space, even in developed turbulence states) is still under analysis, [CL98].

An important prediction of the fluctuation theorem in strongly chaotic particle systems is that the slope of the graph of $\zeta(p) - \zeta(-p)$ is *precisely* σ_+. This shows that even if the distribution of p was Gaussian, *i.e.* $\zeta(p) = \frac{1}{2D}(1-p)^2$ for some $D > 0$, the theorem would be nontrivial. In fact it would be $\zeta(p) - \zeta(-p) = \frac{1}{D}p$ and there would be no *a priori* reason to have $\frac{1}{D} = \sigma_+$.

In general one expects the distribution $\zeta(p)$ to be Gaussian near the average value of p (which is 1 by definition), [Si77], but the Gaussian approximation should be correct only for $|p-1| = O(\tau^{-1/2})$ ("central limit theorem"). Hence it becomes important to test not only the linearity of $\zeta(p) - \zeta(-p)$ but also the slope of this linear law and whether the distribution $\zeta(p)$ can be regarded as Gaussian.

One finds in the first two experiments the correct value of the slope but one cannot really distinguish whether the distribution of p is Gaussian or not. Although one can see *a priori* that it is not Gaussian, the non-Gaussian nature of the distribution is not observable because it manifests itself in a region so far away from $p = 1$ that the corresponding huge fluctuations cannot be observed, being too rare. The attempt at understanding the relation between the central limit theorem and the large deviation experimental results on the fluctuation theorem led to the idea that there was a relation between the fluctuation theorem and the linear response theory of Onsager and Green-Kubo. This in fact was found in [Ga96a],[Ga96b],[Ga98d].

In the third experiment, [LLP97], one finds the correct value of the slope in a situation in which $\zeta(p)$ is manifestly not Gaussian. Hence this is a key experiment for the theory.

Finally in the fourth experiment, [CL98], one gets a linear graph for the odd part of the large deviation function $\zeta(p)$, but the slope is not that of (9.7.12) but considerably smaller. The system in this case, unlike the previous one, is certainly so dissipative that the attractor is much smaller than phase space (the space of the temperature and velocity fields of the sample of water) and the slope was certainly not expected to be σ_+, [BGG97], [BG97]. This might be due to the fact that the system is not reversible, or that it is not equivalent to a reversible one, or that the chaotic hypothesis is incorrect in this case. But the matter requires further investigation, because in the earlier work [Ga96c], [Ga97a] it was shown that in such cases one could expect a slope $\overline{P} < 1$.

A final comment on the observability of the fluctuation theorem in large systems: since the function $\zeta(p)$ is expected to be proportional to the volume of the system, or at least to the surface of its container (depending on the size of the region where dissipation really occurs), it is impossible to observe the fluctuation relation in macroscopic systems because the fluctuations have too small a probability. However *in some cases* it is possible to derive a

"*local fluctuation theorem"* which concerns the fluctuations of the entropy creation rate in a *microscopic* region. In such cases the fluctuations are observable. One is in a situation similar to that of density fluctuations in equilibrium. One cannot see density fluctuations of a gas in a large macroscopic container, but one can quite easily see density fluctuations in a small microscopic volume, and the functions ζ that control such deviations are simply proportional (their ratio being the ratio of the corresponding volumes). It would be interesting to formulate local fluctuation theorems as generally as possible, beyond the few examples known, [Ga98e].

§**9.8. Fluctuation Patterns**

It is natural to inquire whether there are more direct and physical interpretations of the theorem (hence of the meaning of the chaotic hypothesis) when the external forcing is really different from the value 0. A result in this direction is the *conditional reversibility theorem*, discussed below.

Consider an observable F which, for simplicity, has a well-defined time reversal parity: $F(Ix) = \varepsilon_F F(x)$, with $\varepsilon_F = \pm 1$. For simplicity suppose that its time average (*i.e.* its SRB average) vanishes, $F_+ = 0$, and let $t \to \varphi(t)$ be a smooth function vanishing for $|t|$ large enough. We look at the probability, relative to the SRB distribution (*i.e.* in the "natural stationary state") that $F(S_t x)$ is close to $\varphi(t)$ for $t \in [-\frac{\tau}{2}, \frac{\tau}{2}]$. We say that F "follows the fluctuation pattern" φ in the time interval $t \in [-\frac{\tau}{2}, \frac{\tau}{2}]$.

No assumption on the fluctuation size (*i.e.* on the size of φ), nor on the size of the forces keeping the system out of equilibrium, will be made. Besides the chaotic hypothesis we assume, however, that the evolution is *time reversible also out of equilibrium* and that the phase space contraction rate σ_+ is not zero (the results hold no matter how small σ_+ is; and they make sense even if $\sigma_+ = 0$, but they become trivial).

We denote by $\zeta(p,\varphi)$ the *large deviation function* for observing in the time interval $[-\frac{\tau}{2}, \frac{\tau}{2}]$ an average contraction of phase space $\sigma_\tau \stackrel{def}{=} \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} \sigma(S_t x)dt = p\sigma_+$ and at the same time a fluctuation pattern $F(S_t x) = \varphi(t)$.

This means that the probability that the *dimensionless average entropy creation rate* p is in an interval $\Delta = (a,b)$ and, at the same time, F is in a neighborhood[14] $U_{\underline{\psi},\eta}$ of φ, is given by

9.8.1
$$\sup_{p\in\Delta,\varphi\in U_{\underline{\psi},\eta}} e^{-\tau\zeta_\tau(p,\varphi)} \qquad (9.8.1)$$

to leading order as $\tau \to \infty$ (*i.e.* the logarithm of the mentioned probability divided by τ converges as $\tau \to \infty$ to $\sup_{p\in\Delta,\varphi\in U_{\underline{\psi},\varepsilon}} \zeta(p,\varphi)$).

---

[14] By "neighborhood" $U_{\underline{\psi},\eta}$ we mean that $\int_{-\tau/2}^{\tau/2} \psi(t) F(S_t x)dt$ is approximated within given $\eta > 0$ by $\int_{-\tau/2}^{\tau/2} \psi(t)\varphi(t)dt$ for ψ in the finite collection $\underline{\psi} = (\psi_1, \ldots, \psi_m)$ of test functions. This is, essentially, what is called in mathematics a "weak neighborhood".

Given a reversible, dissipative, mixing Anosov flow the fluctuation pattern $t \to \varphi(t)$ and the time reversed pattern $t \to \varepsilon_F \varphi(-t)$ are then related by the following:

Conditional reversibility theorem: Consider a function $t \to \varphi(t)$ and an observable F with defined time reversal parity $\varepsilon_F = \pm 1$. Let τ be large and consider the fluctuation pattern $\{\varphi(t)\}_{t\in[-\frac{\tau}{2},\frac{\tau}{2}]}$ and its time reversal $\{I\varphi(t)\}_{t\in[-\frac{\tau}{2},\frac{\tau}{2}]} \equiv \{\varepsilon_F \varphi(-t)\}_{t\in[-\frac{\tau}{2},\frac{\tau}{2}]}$; they will be followed with equal likelihood if the first is conditioned to an entropy creation rate p and the second to the opposite $-p$. This is an interpretation of the following result:

9.8.2
$$\zeta(-p, I\varphi) = \zeta(p,\varphi) + p\,\sigma_+ \qquad \text{for } |p| < p^* \tag{9.8.2}$$

with ζ introduced above and a suitable $p^* \ge 1$.

In other words, in these systems, while it is very difficult to see an "anomalous" average entropy creation rate during a time τ (*e.g.* $p = -1$), it is also true that "*that is the hardest thing to see*". Once we see it *all the observables will behave strangely* and the relative probabilities of time reversed patterns will become as likely as those of the corresponding direct patterns under "normal" average entropy creation regime.

A waterfall will go up, as likely as we see it going down, in a world in which for some reason, or by the deed of a Daemon, the entropy creation rate has changed sign during a long enough time. We can also say that the motion on an attractor is reversible, even in the presence of dissipation, once the dissipation is fixed.

The proof of the above theorem is similar to that of the fluctuation theorem to which it reduces if $F = \varphi = 0$ (and in fact it is a repetition of it). To be complete we sketch, in the next section, the proof.

The fluctuation and the conditional reversibility theorems can also be formulated for systems whose evolution is studied in continuous time (*i.e.* for Anosov flows). The discrete case is simpler to study than the corresponding Anosov flows because Anosov maps do not have a trivial Lyapunov exponent (the vanishing one associated with the phase space flow direction); the techniques to extend the analysis to Anosov flows are developed in [BR75], [Ge98] (and one achieves the goal of proving the analogue of the fluctuation theorem for such systems).

§9.9. "Conditional Reversibility" and "Fluctuation Theorems"

In §9.4 we have seen that in a Anosov system the stable and unstable tangent planes T^s_x, T^u_x form an integrable family of planes (and their integral surfaces are the stable and unstable manifolds). If x is a point and if $J(x) = \partial S(x)$ is the Jacobian matrix of S at x, then the covariance of the stable and unstable planes implies that we can regard its action (mapping the tangent plane T_x onto T_{Sx}) as "split" linearly into an action on the stable plane and one on the unstable plane: *i.e.* $J(x)$ restricted to the stable plane becomes

a linear map $J^s(x)$ mapping T^s_x to T^s_{Sx}. Likewise one can define the map $J^u(x)$, [Ru79].

Let $\Lambda_u(x), \Lambda_s(x)$ be the determinants of the Jacobians, *i.e.* of $J^u(x)$, $J^s(x)$. Their product differs from the determinant $\Lambda(x)$ of $\partial S(x)$ by the ratio of the sine of the angle $a(x)$ between the planes T^s_x, T^u_x and the sine of the angle $a(Sx)$ between T^s_{Sx}, T^u_{Sx}.

Hence

9.9.1
$$\Lambda(x) = \frac{\sin a(Sx)}{\sin a(x)} \Lambda_s(x) \Lambda_u(x) \tag{9.9.1}$$

and we also set, see §9.5,

9.9.2
$$\begin{aligned} \Lambda_{u,\tau}(x) &= \prod_{j=-\tau/2}^{\tau/2-1} \Lambda_u(S^j x), \quad \Lambda_{s,\tau}(x) = \prod_{j=-\tau/2}^{\tau/2-1} \Lambda_s(S^j x), \\ \Lambda_\tau(x) &= \prod_{j=-\tau/2}^{\tau/2-1} \Lambda(S^j x) \,. \end{aligned} \tag{9.9.2}$$

Time reversal symmetry, which we assume here, implies that $W^s_{Ix} = IW^u_x, W^u_{Ix} = IW^s_x$ and:

9.9.3
$$\begin{aligned} &\Lambda_\tau(x) = \Lambda_\tau(Ix)^{-1}, \quad \Lambda_{s,\tau}(Ix) = \Lambda_{u,\tau}(x)^{-1}, \quad \Lambda_{u,\tau}(Ix) = \Lambda_{s,\tau}(x)^{-1} \\ &\sin a(x) = \sin a(Ix) \,. \end{aligned} \tag{9.9.3}$$

In §9.5 we have seen that, given the above geometric-kinematical notions, the SRB distribution μ can be represented by assigning suitable weights to small phase space cells, (9.6.2). This is very similar to the representation of the Maxwell-Boltzmann distributions of equilibrium states in terms of suitable weights given to phase space cells of equal Liouville volume.

The phase space cells can be made, see §9.5, consistently as small as we please and, by taking them small enough, one can achieve an arbitrary precision in the description of the SRB distribution μ, in the same way as we can approximate the Liouville volume by taking the phase space cells small.

The key to the construction and to our proof is a Markov partition, introduced in §9.5: this is a partition $\mathcal{E} = (E_1, \ldots, E_{\mathcal{N}})$ of the phase space $\mathcal{C}$ into $\mathcal{N}$ cells which are *covariant* with respect to the time evolution and with respect to time reversal in the sense that $IE_j = E_{j'}$ for some j', see §9.5 for the notion of covariance and for the properties of Markov partitions.

Given a Markov partition $\mathcal{E}$ we can "refine" it "consistently", see §9.6, as much as we wish by considering the partition $\mathcal{E}_{-T,T} = \vee_{-T}^{T} S^{-j} \mathcal{E}$ whose cells are obtained by "intersecting" the cells of $\mathcal{E}$ and of its S iterates; the cells of $\mathcal{E}_{-T,T}$ become exponentially small with $T \to \infty$ as a consequence of the hyperbolicity. In each $E_j \in \mathcal{E}_{-T,T}$ one can select a center point $x_j = c(E_j)$ (associated with an arbitrary boundary condition in the sense

of remark (6) in §9.6, see also (9.6.1)), so that Ix_j is the point selected in IE_j. Then we evaluate the expansion rate $\Lambda_{u,2T}(x_j)$ of S^{2T} as a map of the unstable manifold of $S^{-T}x_j$ to that of S^Tx_j, see (9.6.4).

Using the elements $E_j \in \mathcal{E}_T$ as cells we can define approximations "as good as we wish" to the SRB distribution μ, as given by (9.6.4), because for all smooth observables F defined on $\mathcal{C}$, [Si68],[BR75],

9.9.4
$$\int \mu(dy)F(y) = \lim_{T\to\infty} \int m_T(dy)F(y) \stackrel{def}{=} \lim_{T\to\infty} \frac{\sum_{E_j\in\mathcal{E}_T} F(x_j)\Lambda^{-1}_{u,2T}(x_j)}{\sum_{E_j\in\mathcal{E}_T} \Lambda^{-1}_{u,2T}(x_j)} \tag{9.9.4}$$

where $m_T(dy)$ is implicitly defined here by the ratio in the *righ-hand side* of (9.9.4); (9.9.4) is just a change of notation away from (9.6.4).

Let Δ_p denote an interval $[p, p+dp]$ and U_η denote a set of of functions $t \to \varphi'(t)$ defined in $[-\frac{\tau}{2}, \frac{\tau}{2}]$ and with values in a "tube" of width η around a given "path" $t \to \varphi(t)$, $t \in [-\frac{1}{2}\tau, \frac{1}{2}\tau]$. Let εU be the time reversed set of paths, *i.e.* the set of paths $t \to \varphi'(t)$, $t \in [-\frac{1}{2}\tau, \frac{1}{2}\tau]$ with values within η of $\varepsilon_F\varphi(-t)$. Here time is discrete (but the same ideas and deductions would apply to a continuous time case, so that we use a notation that makes sense in both cases.

We *first evaluate the probability*, with respect to the distribution $m_{\tau/2}$ in (9.9.4), (*instead of the* m_T), of the event that $p = p(x_j) = \sigma_\tau(x_j)/\langle\sigma\rangle_\mu \in \Delta_p$ and, also, that $\{F(S^kx_j)\}_{k=-\tau}^{\tau} \in U_\eta$, divided by the probability (with respect to the same distribution) of the time reversed event that $p(x_j) = \sigma_\tau(x_j)/\langle\sigma\rangle_+ \in \Delta_{-p}$ and, also, $k \to \{F(S^kx_j)\} \in \varepsilon U_\eta$.

Thus we compare the probability of a fluctuation pattern φ in the presence of average dissipation p and that of the time reversed pattern in the presence of average dissipation $-p$. This is essentially:

9.9.5
$$\frac{\pi_\tau(p)}{\pi_\tau(-p)} = \frac{\sum_{j,\,p(x_j)=p,F(S^nx_j)=\varphi(n)} \Lambda^{-1}_{u,\tau}(x_j)}{\sum_{j,\,p(x_j)=-p,F(S^nx_j)=\varepsilon_F\varphi(-n)} \Lambda^{-1}_{u,\tau}(x_j)} \tag{9.9.5}$$

Since $m_{\tau/2}$ in (9.9.4) is only an approximation to μ_+ *an error is involved in using* (9.9.5) as a formula for the same ratio computed by using the true SRB distribution μ instead of $m_{\tau/2}$.

It can be shown that this "first" approximation (among the two that will be made) can be estimated to affect the result only by a factor bounded above and below uniformly in τ, p. This is *not* completely straightforward: in a sense this is perhaps the main technical problem of the analysis.**[15]** Further mathematical details can be found in [Ga95c],[Ru97c],[Ge98].

---

15 It can be seen if one interprets (9.9.5) as a probability distribution on the space of the symbolic sequences $\underline{\sigma}$ which, via the Markov partition $\mathcal{E}$, can be used to represent the points x in phase space. Such probability distribution can *also* be interpreted as a Gibbs distribution over the space of the sequences $\underline{\sigma}$ with potential $\lambda(\underline{\sigma}) = \log\Lambda_{u,1}(x(\underline{\sigma}))$, if $\underline{\sigma}$ is the symbolic sequence corresponding to x_j: see §5.10. In this way the property under analysis (*i.e.* the identity of the limits as $\tau \to \infty$ of (9.9.5) and of the same ratios evaluated by using m_T instead of $m_{[-\frac{1}{2}\tau,\frac{1}{2}\tau]}$), appears simply due to the nonexistence

Remark: There are other representations of the SRB distributions that seem more appealing than the above one based on the Markov partitions notion and still make the above analysis possible and apparently more intuitive, *e.g.* see [MR97]. The simplest is perhaps the *periodic orbits representation* in which the role of the cells is taken by the periodic orbits. However I do not know a way of making the argument that leads to (9.9.5) while *keeping under control* the approximations and at the same time not relying on Markov partitions; and in fact I do not know of *any expression* of the SRB distribution that is not proved by using the very existence of Markov partitions.

We now try to establish a one-to-one correspondence between the addends in the numerator of (9.9.5) and those in the denominator, aiming at showing that corresponding addends have a *constant ratio* which will, therefore, be the value of the ratio in (9.9.5).

This is possible because of the reversibility property, it will be used in the form of its consequences given by the relations (9.9.3). The ratio (9.9.5) can therefore be written, by virtue of (9.9.3), simply as:

9.9.6
$$\frac{\sum_{j,\,p(x_j)=p,F(S^n x_j)=\varphi(n)}\Lambda^{-1}_{u,\tau}(x_j)}{\sum_{j,\,p(x_j)=-p,F(S^n x_j)=\varepsilon_F\varphi(-n)}\Lambda^{-1}_{u,\tau}(x_j)}\equiv\frac{\sum_{j,\,p(x_j)=p,F(S^n x_j)=\varphi(n)}\Lambda^{-1}_{u,\tau}(x_j)}{\sum_{j,\,p(x_j)=p,F(S^n x_j)=\varphi(n)}\Lambda_{s,\tau}(x_j)} \tag{9.9.6}$$

where $x_j\in E_j$ is the center in E_j. In deducing the second relation we take into account:

(1) time reversal symmetry I,

(2) that the centers $x_j, x_{j'}$ of E_j and $E_{j'}=IE_j$ are such that $x_{j'}=Ix_j$, and

(3) that (9.9.3), (9.9.2) hold,

and transform the sum in the denominator of the left-hand side of (9.9.6) into a sum over the same set of labels that appear in the numerator sum.

It follows then that the ratios between corresponding terms in the ratio (9.9.6) is equal to $\Lambda^{-1}_{u,\tau}(x)\Lambda^{-1}_{s,\tau}(x)$. This *differs little* from the reciprocal of the total change of phase space volume over the τ time steps (during which the system evolves from the point $S^{-\tau/2}x$ to $S^{\tau/2}x$).

The difference is only due to not taking into account the ratio of the sines of the angles $a(S^{-\tau/2}x_j)$ and, see (9.9.2), $a(S^{\tau/2}x_j)$ formed by the stable and unstable manifolds at the points $S^{-\tau/2}x_j$ and $S^{\tau/2}x_j$. Therefore

---

of phase transitions in the one-dimensional short-range Ising models. In fact the two ratios become ratios of expectation values of the same quantities evaluated in presence of different boundary conditions, and in absence of phase transitions one should have boundary conditions independence which in this case would imply that the two ratios differ at most by a factor of order $O(1)$ so that their logarithms divided by T or $\frac{\tau}{2}$ should have the *same* limit $\zeta(p)-\zeta(-p)$.

$\Lambda_{u,\tau}^{-1}(x_j)\Lambda_{s,\tau}^{-1}(x_j)$ will differ from the actual phase space contraction under the action of S^τ, regarded as a map between $S^{-\tau/2}x_j$ and $S^{\tau/2}x_j$, by a factor that can be bounded between B^{-1} and B with $B = \max_{x,x'}\big(\frac{|\sin a(x)|}{|\sin a(x')|}\big)$, which is finite and positive, by the linear independence of the stable and unstable planes.

But for all points x_j in (9.9.6), the reciprocal of the total phase space volume change over a time τ equals $e^{\tau\sigma_+ p}$ (*by the constraint,* $\sigma_\tau/\sigma_+ = p$, *imposed on the summation labels*) up to a "second" approximation that cannot exceed a factor which is bounded above and below by τ*-independent* positive and finite constants $B^{\pm1}$, due to the above sine ratio. Hence the ratio (9.9.5) will be the exponential $e^{\tau\sigma_+ p}$, up to a τ-independently bounded factor and (9.9.3) follows.

It is important to note that there have been two approximations, as just pointed out. They can be estimated, see [GC95],[Ga95c], [Ru97c], and imply that the argument of the exponential *is correct up to* p,φ,τ*-independent corrections* so that the result can be proved even if the approximations are avoided (this also makes clear that the consideration of the limit $\tau\to\infty$ is *necessary* for the theorem to hold).

In the special cases in which there is no F and one only looks at the probability distribution of the entropy production rate the above becomes

9.9.7
$$B^{-1}\, e^{\tau\, p\, \langle\sigma\rangle_+} < \frac{\pi_\tau(p)}{\pi_\tau(-p)} < B\, e^{\tau\, p\, \langle\sigma\rangle_+} \tag{9.9.7}$$

or

9.9.8
$$\frac{\pi_\tau(p)}{\pi_\tau(-p)} = e^{\tau\, p\, \langle\sigma\rangle_+ + O(1)} \tag{9.9.8}$$

i.e. we get (9.7.12).

§9.10. Onsager Reciprocity and Green-Kubo's Formula.

The fluctuation theorem degenerates in the limit in which σ_+ tends to zero, *i.e.* when the external forces vanish and dissipation disappears (and the stationary state becomes the equilibrium state).

Since the theorem deals with systems that are time reversible *at and outside* equilibrium, Onsager's hypotheses are certainly satisfied and the system should obey reciprocal response relations at vanishing forcing. This led to the idea that there might be a connection between the fluctuation theorem and Onsager reciprocity and also to the related (stronger) Green-Kubo formula.

This is in fact true: if we define the *microscopic thermodynamic flux* $j(x)$ associated with the *thermodynamic force* E that generates it, *i.e.* the parameter that measures the strength of the forcing (which makes the system nonHamiltonian), via the relation

9.10.1
$$j(x) = \frac{\partial\sigma(x)}{\partial E} \tag{9.10.1}$$

(not necessarily at $E=0$) then in [Ga96b] a heuristic proof shows that the limit as $E\to0$ of the fluctuation theorem becomes simply (in the continuous time case) a property of the average, or "macroscopic", *flux* $J=\langle j\rangle_{\mu_E}$:

9.10.2
$$\frac{\partial J}{\partial E}\Big|_{E=0}=\frac12\int_{-\infty}^{\infty}\langle j(S_tx)j(x)\rangle_{\mu_E}\Big|_{E=0}dt \tag{9.10.2}$$

where $\langle\cdot\rangle_{\mu_E}$ denotes the average in the stationary state μ_E (*i.e.* the SRB distribution which, at $E=0$, is simply the microcanonical ensemble).

If there are several fields $E_1,E_2,\ldots$ acting on the system we can define several thermodynamic fluxes $j_k(x)\stackrel{def}{=}\partial_{E_k}\sigma(x)$ and their averages $\langle j_k\rangle_\mu$: in the limit in which all forces E_k vanish a (simple) extension of the fluctuation theorem is shown, [Ga96b], to reduce to

9.10.3
$$L_{hk}\stackrel{def}{=}\frac{\partial J_h}{\partial E_k}\Big|_{E=0}=\frac12\int_{-\infty}^{\infty}\langle j_h(S_tx)j_k(x)\rangle_{E=0}\,dt=L_{kh}\,, \tag{9.10.3}$$

therefore we see that the fluctuation theorem can be regarded as *an extension to nonzero forcing* of Onsager reciprocity and, actually, of the Green-Kubo formula.

Certainly assuming reversibility in a system out of equilibrium can be disturbing: one can, thus, inquire if there is a more general connection between the chaotic hypothesis, Onsager reciprocity and the Green-Kubo formula.

This is indeed the case and provides us with a further consequence of the chaotic hypothesis valid, however, only in zero field. It can be shown that the relations (9.10.3) follow from the sole assumption that at $E=0$ the system is time reversible and that it satisfies the chaotic hypothesis for E near 0: at $E\ne0$ it can be, as in Onsager's theory, not reversible [GR97].

It is not difficult to see, technically, how the fluctuation theorem, in the limit in which the driving forces tend to 0, formally yields the Green-Kubo formula.

We consider time evolution in continuous time and simply note that (9.9.8) implies that, for all E (for which the system is chaotic) $\langle e^{I_E}\rangle=\sum_p\pi_\tau(p)e^{p\tau\sigma_+}=\sum_p\pi_t(-p)e^{O(1)}=e^{O(1)}$ so that:

9.10.4
$$\lim_{\tau\to+\infty}\frac1\tau\log\langle e^{I_E}\rangle_{\mu_E}=0 \tag{9.10.4}$$

where $I_E\stackrel{def}{=}\int\sigma(S_tx)dt$ with $\sigma(x)$ being the divergence of the equations of motion (*i.e.* the phase space contraction rate, in the case of continuous time). This remark, [Bo97b],[16] can be used to simplify the analysis in [Ga96b] (and [Ga96a]) as follows.

We switch to continuous time, to simplify the analysis. Differentiating both sides with respect to E, not worrying about interchanging derivatives and

[16] It says that essentially $\langle e^{I_E}\rangle_{\mu_E}\equiv1$ or more precisely it is not too far from 1 so that (9.10.4) holds.

limits and the like, one finds that the second derivative with respect to E is a sum of six terms. Supposing that for $E = 0$ the system is Hamiltonian and (hence) $I_0 \equiv 0$, the six terms, when evaluated at $E = 0$, are:

9.10.5
$$\begin{aligned}&\frac{1}{\tau}\Big[\,\langle\partial_E^2 I_E\rangle_{\mu_E}|_{E=0} - \langle(\partial_E I_E)^2\rangle_{\mu_E}|_{E=0}+\\ &+\int \partial_E I_E(x)\partial\mu_E(x)|_{E=0} - \Big(\langle(\partial_E I_E)^2\rangle_{\mu_E}\cdot\int 1\,\partial_E\mu_E\Big)|_{E=0}+\\ &+\int \partial_E I_E(x)\partial\mu_E(x)|_{E=0} + \int 1\cdot\partial_E^2\mu_E|_{E=0}\,\Big]\end{aligned} \tag{9.10.5}$$

and we see that the fourth and sixth terms vanish being derivatives of $\int \mu_E(dx) \equiv 1$, and the first vanishes (by integration by parts) because I_E is a divergence and μ_0 is the Liouville distribution (by the assumption that the system is Hamiltonian at $E = 0$ and chaotic). Hence we are left with:

9.10.6
$$\Big(-\frac{1}{\tau}\langle(\partial_E I_E)^2\rangle_{\mu_E} + \frac{2}{\tau}\int \partial_E I_E(x)\partial_E\mu_E(x)\Big)_{E=0} = 0 \tag{9.10.6}$$

where the second term is $2\tau^{-1}\partial_E(\langle\partial_E I_E\rangle_{\mu_E})|_{E=0} \equiv 2\partial_E J_E|_{E=0}$, because the distribution μ_E is stationary; and the first term tends to $\int_{-\infty}^{+\infty}\langle j(S_t x)j(x)\rangle_{E=0}dt$ as $\tau \to \infty$. Hence we get the Green-Kubo formula in the case of only one forcing parameter.

The argument could be extended to the case in which E is a vector describing the strength of various driving forces acting on the system, but one needs a generalization of (9.10.4). The latter is a consequence of the fluctuation theorem, and the theorem has to be extended in order to derive from it also the Green-Kubo formula (hence reciprocity) when there were several independent forces acting on the system; see [Ga96b] where the extension is discussed.

The above analysis is unsatisfactory because we interchange limits and derivatives quite freely and we even take derivatives of μ_E, which seems to require some imagination as μ_E is concentrated on a set of zero volume. On the other hand, under the strong hypotheses in which we suppose to be working (that the system is mixing Anosov), we should not need extra assumptions. *Indeed* the above mentioned nonheuristic analysis, [GR97], is based on the solution of the problem of differentiability with respect to a parameter for SRB distributions, [Ru97b].

§**9.11. Reversible Versus Irreversible Dissipation. Nonequilibrium Ensembles?**

What is missing are arguments similar to those used by Boltzmann to justify the use of the ensembles *independently* of the ergodic hypothesis: an hypothesis which in the end may appear (and still does appear to many) as having led to the theory of enembles only “by accident”. The missing arguments should justify the fluctuation theorem on the basis of the extreme

likelihood of its predictions in systems that are very large and that may not be Anosov systems in the mathematical sense. I see no reason why this should prove impossible, *a priori*, now or in the future.

In the meantime it seems interesting to take the same philosophical attitude adopted by Boltzmann: not to consider that "by chance" *chaotic systems share some selected properties*, and try to see if such properties help us achieve a better understanding of nonequilibrium. After all it seems that Boltzmann himself took a rather long time to realize the interplay of the above two basic mechanisms behind the equilibrium ensembles and to propose a solution harmonizing them. "All it remains to do" is to explore if the hypothesis has implications more interesting or deeper than the fluctuation theorem.

A system driven out of equilibrium can reach a stationary state (and not steam out of sight) only if enough dissipation is present. This means that any mechanical model of a system reaching a stationary state out of equilibrium *must* be a model with nonconservative equations of motion in which forces representing the action of the thermostats, that keep the system from heating up, are present.

Thus, as we stressed repeatedly in the previous sections, a generic model of a system stationarily driven out of equilibrium will be obtained by adding to Hamilton's equations (corresponding to the nondriven system) other terms representing forces due to the thermostat action.

Here one should avoid attributing a fundamental role to special assumptions about such forces. One has to realize that there is *no privileged* thermostat: many of them can be considered and they simply describe various ways to take energy out of the system.

Hence one can even use stochastic thermostats, and there are many types considered in the literature; or one can consider deterministic thermostats and, among them, reversible ones or irreversible ones.

Each thermostat requires its own theory. However the same system may behave in the same way under the action of different thermostatting mechanisms: if the only action we make on a gas tube is to keep the temperatures of its extremes fixed, by taking in or out heat from them, the difference may be irrelevant, at least in the limit in which the tube becomes long enough and as far as what happens in the middle of it is concerned.

But of course the mathematical representation of the stationary state may be very different in the various cases, even when we think that the differences are only minor boundary effects.

For instance, in the case of the gas tube, if our model is of deterministic dissipation we expect the SRB state to be concentrated on a set of *zero phase space volume*,[17] while if the model is stochastic then the stationary state will be described by a *density* on phase space. Nothing could seem more different.

---

[17] Because phase space will on the average contract, when $\sigma_+ > 0$, so that any stationary state has to be concentrated on a set of zero volume, *which however could still be dense* and often it will be.

Nevertheless it might still be true that in the limit of an infinite tube the two models give the same result, in the same sense as the canonical and microcanonical ensembles describe the same state even though the microcanonical ensemble is supported on the energy surface, which has zero volume if measured by using the canonical ensemble (which is given by a density over the whole available phase space).

Therefore we see that out of equilibrium we have in fact *much more freedom to define equivalent ensembles.* Not only do we have (very likely) the same freedom that we have in equilibrium (like fixing the total energy or not, or fixing the number of particles or not, passing from microcanonical to canonical to grand canonical, *etc*) but *we can also change the equations of motion and obtain different stationary states, i.e. different SRB distributions, which will however become the same in the thermodynamic limit.*

Being able to prove mathematical equivalence of two thermostats will amount to proving their physical equivalence. This again will be a difficult task, in any concrete case.

What I find fascinating is that the above remarks seem to indicate to us the possibility that a *reversible thermostat can be equivalent in the thermodynamic limit to an irreversible one.* I conclude by reformulating a conjecture, see for instance [Ga96c], [Ga97a], [Ga98c], which clarifies the latter statement.

Consider the following two models describing a system of hard balls in a periodic (large) box in which there is a lattice of obstacles that forbid collisionless paths (by their arrangement and size); the laws of motion will be Newton's laws (elastic collisions with the obstacles as well as between particles) plus a constant force E along the horizontal axis (say) *plus a thermostatting force.*

In the first model the thermostatting force is simply a constant times the momentum of the particles: it acts on the i-th particle as $-\nu p_i$ if ν is a "friction" constant. Another model is a force proportional to the momentum but via a proportionality factor that is not constant and depends on the system configuration at the point x in phase space; it has the form $-\alpha(x)p_i$ with $\alpha(x) = E \cdot \sum_i p_i / \sum_i p_i^2$.

The first model is related to the model used by Drude in his theory of conduction in metals, see [EGM98]. The second model has been used very often in recent years for theoretical studies and has thus acquired a "respected" status and a special importance: it was among the first models used in the experiments and theoretical ideas that led to the connection between Ruelle's ideas for turbulent motion in fluids and nonequilibrium statistical mechanics, [HHP87], [ECM90], [ECM93]. I think that the importance of such work should be stressed and fully appreciated: without this work the recent theoretical developments would have been unthinkable, in spite of the fact that *a posteriori* they seem quite independent and one could claim (unreasonably in my view) that everything could have been done much earlier.

Furthermore the second model can be seen as derived from Gauss' least constraint principle, see Appendix 9.A4. It keeps the total (kinetic) energy exactly constant over time (taking energy in and out, as needed) and is called a *Gaussian thermostat. Unlike the first model the second model is reversible*, with time reversal being the usual velocity inversion. Thus the above theory and results based on the chaotic hypothesis apply.

The conjecture was (and is) that:

(1) Compute the average energy per particle that the system has in the constant friction case and call it $\mathcal{E}(\nu)$ calling also μ_ν the corresponding SRB distribution.

(2) Call $\tilde\mu_{\mathcal{E}}$ the SRB distribution for the Gaussian thermostat system when the total (kinetic) energy is fixed to the value $\mathcal{E}$.

(3) Then $\mu_\nu = \tilde\mu_{\mathcal{E}(\nu)}$ *in the thermodynamic limit* (in which the box size tends to become infinitely large, with the number of particles and the total energy correspondingly growing so that one keeps the density and the energy density constant) and for *local* observables, *i.e.* for observables that depend only on the particles of the system localized in a fixed finite region of the container. This means that the equality takes place in the usual sense of the theory of ensembles, see Chap.II,IV and [Ru68].

It has to be remarked that the idea of equivalence between dynamical ensembles, in contexts perhaps more limited, seems to circulate for quite a long time particularly among those who work on numerical experiments: remarkable are the early papers [ES93], [Ev93], [SJ93] which certainly propose the same kind of ideas, see also [MR96].

The above conjecture opens the way to several speculations as it shows that the reversibility assumption might be not so strong after all. And results for reversible systems may carry through to irreversible ones.

I have attempted to extend the above ideas also to cases of turbulent motions but here I can only give references, [Ga97a],[Ga97b].

There are a few other results and many speculations about the consequences of the chaotic hypothesis: among the (few) related results. I want to quote the "*pairing rule*", valid for a somewhat restricted class of systems. It is a further extremely interesting example of a mathematical theorem discovered through physical experiments and, although heralded by a similar result in a simpler case, [Dr88], it was proved only later (like the fluctuation theorem), [ECM90], [EM90], [DM96], [WL98]. It is related to the chaotic hypothesis but *it does not depend on it*, the relation being that it emerged in the same group of experiments that led to the fluctuation theorem.

Among the "speculations" I quote:

(1) Several applications to fluid mechanics, like the equivalence of Navier-Stokes equations to a similar *reversible*, equation (in the limit of large Reynolds' number), [Ga96c], [Ga97a], [SJ93].

(2) stability of time reversal symmetry whereby under assumptions, that I think are quite natural, [BG97], one deduces that when time reversal is spontaneously broken[18] it is replaced by another symmetry with the same property of anticommuting with time evolution, thus showing that the fluctuation theorem might hold, with minor modifications, even in cases in which the time reversal symmetry of the equations of motion is broken, see §1.6 of [BGG97].

(3) The possibility of equivalence between reversible and irreversible equations describing the same system, in the limit of large systems, [Ga98c]. This gives hope that some result like a, suitably reformulated, fluctuation theorem can hold even in irreversibly driven systems. Furthermore it seems to indicate that the theory of ensembles in nonequilibrium is much richer than what we are used to in equilibrium. An ensemble might be characterized not only by the choice of a few parameters, as in Chap.II, but *also* by the choice of the equations of motion.

The above incomplete list is here only to provide the reader with a guide to the literature, which is constantly increasing but which does not yet seem established enough to be treated as an accomplished theory deserving more space in a short treatise.

Appendix 9.A1. Mécanique statistique hors équilibre: l'héritage de Boltzmann

The following is a slightly expanded version of a talk at École Normale Superieure in Paris, january 1998, and gives an informal overview of the basic ideas of Chap.IX, see §9.3, and some supplementary analysis.

Boltzmann entreprit, [Bo66], de prouver l'existence des atomes en poursuivant un programme déjà amorcé par ses prédécesseurs. Son approche était d'établir que la conception de la matière en tant qu'agglomération d'atomes obéissants aux lois de la mécanique conduisait à la déduction des propriétés de la matière que connaissaient alors les expérimentateurs et les théoriciens.
Ainsi Boltzmann produisit des versions de plus en plus raffinées du *théorème de la chaleur*, [Bo68], [Bo71], [Bo72], [Bo77]. Au début il s'agissait de faire voir qu'il est possible de définir des quantités mécaniques associées, par exemple, à un gaz enfermé dans un conteneur cubique de volume V, telles que :

T = *énergie cinétique moyenne*
U = *énergie totale*
V = *volume*
p = *impulsion moyenne transférée aux parois par collision et par unité de surface*

[18] Because the attracting set A becomes strictly smaller than the phase space and $IA \neq A$.

où les moyennes sont calculées empiriquement en supposant les particules indépendantes et à distribution uniforme sur une sphère dans l'espace des impulsions et dans le volume V des positions.

Le théorème à prouver est alors que si on varie U et V de dU et dV et que l'on calcule la quantité :

9.A1.1
$$\frac{dU + p\,dV}{T} \tag{9.A1.1}$$

où p et T dépendent de U et V, on trouve *une différentielle exacte* : c'est à dire qu'il existe une fonction $S(U,V)$ telle que $dS = \frac{dU + p\,dV}{T}$. Suite aux travaux de Boltzmann, Helmholtz considéra les systèmes mécaniques *monocycliques*, c'est-à-dire les systèmes dont tout mouvement d'énergie donnée est périodique et non dégénéré (ce qui veut dire que les mouvements d'énergie donnée ne diffèrent entre eux que par un décalage du temps d'observation), [He95a],[He95b].

Il fit voir que, en général, si on imagine que les mouvements ("*états*") de tels systèmes sont paramétrés par leur énergie totale U et par un paramètre V dont les potentiels φ_V des forces qui agissent sur le système dépendent, alors en définissant :

T = *énergie cinétique moyenne*
U = *énergie totale*
V = *volume*
$p = \langle -\frac{\partial}{\partial V}\varphi \rangle$

où $\langle F \rangle$ maintenant dénote précisément la moyenne de F par rapport au temps (et donc n'est pas définie empiriquement comme dans le cas précedent), on trouve *en général* :

9.A1.2
$$\frac{dU + p\,dV}{T} = \text{différentielle exacte} \tag{9.A1.2}$$

Celle-ci aurait pu n'être rien de plus qu'une curiosité. Mais Boltzmann avait une conception *discrète* de la nature : même s'il ne l'avait pas explicitement dit dans ses écrits populaires, on le verrait dans ses travaux scientifiques où l'emploi de l'analyse, avec ses intégrales et ses dérivées, est souvent vu comme un moyen technique pour venir à bout du calcul de sommes et de différences, [Bo74].

Donc pour Boltzmann le mouvement n'est qu'une évolution discrète où l'espace des phases est quadrillé en petites cellules à $6N$ dimensions (N étant le nombre de molécules) dont *une* contient le point qui représente l'état instantané du système. L'évolution apparait comme les déplacements successifs du point représentatif d'une cellule à une autre, alors que le temps s'écoule d'une petite quantité discrète h. Bien sûr le déplacement doit être conforme aux lois du mouvement.

C'est une représentation très familière aujourd'hui à qui essaye de simuler sur ordinateur les mouvements d'un gaz de particules. Sur l'ordinateur les états microscopiques du gaz sont représentés par des cellules (car les coordonnées des points sont représentées par des nombres qui sont déterminés avec une précision qui est loin d'être infinie et qui dépend de la

machine ou plutôt du logiciel que l'on emploit) et l'évolution se déroule par pas discrets; le programme qui effectue ces pas est écrit avec les lois du mouvement comme guide.

De ce point de vue le mouvement est une permutation des cellules qui représentent l'état microscopique. Le système est alors toujours en évolution périodique : car toute permutation d'un nombre fini d'objets (les cellules d'énergie totale U donnée, dans le cas présent) engendre une évolution cyclique.

On imagine que l'on fixe l'énergie totale U et que les forces agissantes sur le système sont paramétrées par le volume V : en fait on imagine que, quoi que l'on fasse, les forces entre les particules ne varient pas et seules les forces entre les particules et les parois peuvent changer (à cause des mouvements des parois et des changements de volume qui en découlent).

Alors l'hypothèse de monocyclicité de Helmholtz, de non-dégénérescence des mouvements d'énergie donnée, correspondrait à dire que l'évolution est une permutation à un seul cycle des cellules et donc on serait dans la situation où le système est monocyclique : cette hypothèse est connue comme l'*hypothèse ergodique*.

Sous cette hypothèse on devrait avoir la possiblité de trouver, *en général*, une analogie mécanique de la thermodynamique et un théorème général de la chaleur. Puisque les moyennes doivent se calculer, alors, par la distribution uniforme sur l'espace des cellules d'énergie donnée (car les cellules ont des tailles égales) on se trouve obligé de vérifier que dans le cas d'un gaz (ou même d'un liquide ou d'un solide, vue la généralités des considérations en question):

9.*A*1.3
$$\frac{dU\,+\,p\,dV}{T}\quad\text{est exact si}\quad p=-\langle\frac{\partial}{\partial V}\varphi_V\rangle. \tag{9.A1.3}$$

Cette propriété, cas particulier d'une propriété plus générale que Boltzmann appella *orthodicité*, doit être accompagnée par la propriété supplémentaire que p est *aussi* l'impulsion moyenne transférée aux parois par les collisions, par unité de temps et de surface. Si cela est bien le cas on aura prouvé que en général un théorème de la chaleur est valable.

C'est ce que Boltzmann fit en 1884, [Bo84], en fondant, en même temps, la *théorie des ensembles statistiques* (qui est souvent attribuée à Gibbs, mais pas par Gibbs lui même, [Gi81]).

Il est tout à fait remarquable que le théorème de la chaleur, (9.A1.2), est valable tant pour les petits systèmes (même à une particule, si la non-linéarité du mouvement est suffisante de façon à rendre l'hypothèse ergodique raisonnable) que pour les grands (avec 10^{23} particules).

L'exactitude de $\dfrac{dU\,+\,p\,dV}{T}$ *ne dépend pas de la taille du système*, [Bo84].

Cette indépendance est d'ailleurs une propriété absolument fondamentale et elle permit à Boltzmann de se dégager des critiques qui lui étaient adressées.

Les critiques, par Zermelo et même par Poincaré, étaient subtiles et portaient sur le principe selon lequel il serait impossible de déduire les lois macroscopiques (irréversibles) d'une mécanique réversible qui est nécessairement cyclique et donc apparemment pas irréversible (au bout d'un temps de *récurrence* le système revient à son état initial, contre toute intuition sur le comportement des systèmes macroscopiques), [Bo96], [Bo97].

Ces critiques s'adressaient surtout à l'équation de Boltzmann et par conséquent à l'approche irréversible à l'équilibre. Boltzmann, comme il est bien connu, répondit qu'on ne pouvait pas ne pas tenir compte des échelles de temps nécessaires à réveler des contradictions. Pour voir, au niveau macroscopique, les effets de la réversibilité microscopique, le temps qu'il fallait attendre était énorme qu'on le mesure en heures ou en âges de l'Univers, [Bo74]. Après quoi on observerait une évolution anormale pour revenir presque immédiatement au comportement normal et pour une période de durée encore aussi longue.

Mais cet argument, à la défense de l'équation de Boltzmann, détruisait aussi apparemment la signification du théorème de la chaleur et la possibilité de déduire la thermodynamique de la mécanique et de l'hypothèse ergodique. Car pour que le théorème de la chaleur ait un intérêt quelconque il faut que les moyennes dont il parle soient atteintes dans un laps de temps raisonnablement court : mais si le temps de récurrence (c'est à dire le temps nécessaire au point représentatif du système pour revenir à la cellule initiale dans l'espace des phases) est énorme alors les moyennes des observables risquent d'être atteintes sur un temps du même ordre, ce qui signifierait qu'elles n'ont pas d'intérêt physique.

Boltzmann aperçut cette difficulté et fut conduit à dire que dans un système macroscopique tout se passe comme si les moyennes sur des temps courts étaient les mêmes que sur les temps (inobservables) de récurrence. Ceci serait dû au fait que si le nombre de particules est très grand, les grandeurs d'intérêt thermodynamique prennent la même valeur sur presque tout l'espace des phases : ce qui leur permet d'atteindre leur valeur moyenne sur des temps très courts qui n'ont rien à voir avec le temps de récurrence (qui est infini à tout point de vue). Elle prennent la même valeur parce qu'elles sont à leur tour des moyennes sur les particules et ne dépendent pas de l'état de particules individuelles.

Donc l'hypothèse ergodique suggère l'ensemble microcanonique pour le calcul des moyennes : c'est un fait général que ces moyennes vérifient les relations thermodynamiques qui, d'un autre côté, sont observables grâce à la lois des grands nombres qui fait que ces grandeurs ont la même valeur partout (ou presque) dans l'espace des phases.

Il s'en suit que l'hypothèse ergodique n'est pas une justification de la thermodynamique et ne joue qu'un rôle cinématique. La thermodynamique est une *identité* mécanique qui devient observable au niveau macroscopique grâce à la loi des grands nombres, (§1.9).

Une fois achevée cette admirable construction conceptuelle on se pose la

question de savoir si on peut faire de même dans le cas des systèmes *hors équilibre*.

Ce sont des systèmes de particules sur lesquels agissent une ou plusieurs forces conservatives dont le travail est dissipé dans des thermostats, permettant ainsi au système d'atteindre un état stationnaire.

C'est un problème pas vraiment touché par Boltzmann qui étudia en détail le problème du retour à l'équilibre d'un gaz perturbé de son état d'équilibre (retour qui se déroule selon l'équation de Boltzmann). Et il peut paraître étrange qu'un problème si naturel et d'une telle importance soit resté essentiellement ouvert jusqu'à nos jours.

On remarque immédiatement une profonde différence par rapport au problème de la théorie des états d'équilibre : il n'y a pas une véritable théorie macroscopique (comparable à la thermodynamique classique) qui puisse servir de guide et qui fournisse des résultats à prouver.

Une différence technique importante est que l'on peut s'attendre à ce que le comportement physique du système dépende de la méthode qu'on emploie pour enlever la chaleur produite par le travail des forces qui agissent. Ce qui peut donner le souci qu'une théorie générale soit impossible à cause de la grande variété de forces thermostatiques qu'on peut imaginer pour un même système.

Mais, à mon avis, il ne s'agit que d'une difficulté apparente qui disparait au fur et à mesure qu'on précise la théorie.

Donc on va imaginer un système de particules sur lesquelles agissent des forces externes non conservatives et un mécanisme quelconque qui empèche le réchauffement. On va modéliser ce thermostat par des forces additionnelles. Par exemple, si le système est un gaz de sphères dures enfermées dans un conteneur périodique avec quelques obstacles fixes et soumises à un champ de force $\underline{E}$, on peut imaginer que les équations du mouvement soient :

9.*A*1.4
$$m\underline{\ddot{x}}_i = \underline{f}_i + \underline{E} - \nu\underline{\dot{x}}_i = \Phi_i(\underline{x}, \underline{\dot{x}}) \tag{9.A1.4}$$

où les $\underline{f}_i$ sont les forces entre particules (sphères dures élastiques) et entre particules et obstacles (qui sont aussi des sphères dures élastiques).

Ici $\nu\underline{\dot{x}} = \nu(\underline{\dot{x}})\,\underline{\dot{x}}$ est le modèle de thermostat. La vraie *difficulté* est que l'évolution engendre une contraction du volume de l'espace des phases car :

9.*A*1.5
$$\frac{d}{dt}(d\underline{x}\,d\underline{\dot{x}}) = \mathrm{div}\Phi\,\cdot\,(d\underline{x}\,d\underline{\dot{x}}) \tag{9.A1.5}$$

et la dissipativité entraine $-\langle \mathrm{div}\Phi\rangle > 0$, et donc l'état stationnaire devra être une distribution de probabilité $\mu(d\underline{x}, d\underline{\dot{x}})$ *concentrée sur un ensemble de volume nul*. Elle ne pourra pas être décrite par une densité de la forme : $\rho(\underline{x}, \underline{\dot{x}})\,d\underline{x}, d\underline{\dot{x}}$.

Du coup on ne peut même pas écrire les formules qui expriment formellement les moyennes des observables par rapport à l'état stationnaire en termes d'une fonction de densité inconnue.

Néanmoins on voudrait avoir de telles expressions pour pouvoir espérer en tirer des conséquences générales, du type du théorème de la chaleur, qui puissent être observées dans les petits systèmes (parce que directement observables) et dans les grands aussi (pour des raison différentes).

L'idée clef a pris forme au début des années 1970, 1973 au plus tard, et est due à Ruelle : mais dans un contexte apparemment assez différent du nôtre (celui de la mécanique des fluides et de la turbulence). On conçoit les mouvements turbulents d'un fluide stationnaire ou d'un gaz de particules comme des *mouvements chaotiques*.

Cela ne demande pas à première vue beaucoup d'imagination : mais le point est que l'hypothèse est posée dans un sens technique précis, [Ru78a], [Ru80]. Dans l'interprétation d'auteurs successifs, [GC95], on dit que le principe est que *le système est "hyperbolique" ou d' "Anosov"*. C'est l'*hypothèse chaotique*.

Cela veut dire que en tout point x de l'espace des phases on peut établir un système *covariant* de coordonnées locales tel que l'évolution temporelle $n \to S^n x$ observée dans ce système voit x comme un point fixe (car on le suit) hyperbolique. C'est-à-dire on voit depuis x les autres points bouger de la même façon qu'on les voit si on regarde les mouvements à partir du point fixe instable d'un pendule : la différence étant que cela est vrai pour tout point (et non pas pour un point isolé comme dans le cas du pendule).

On aura cette propriété valable à l'équilibre aussi bien que hors équilibre : les mouvements des molécules sont chaotiques même dans les états d'équilibre. Pour comprendre ce qui se passe il convient de revenir au point de vue discret de Boltzmann.

Si un système est dissipatif il y a des difficultés supplémentaires car il est clair qu'on a beau rendre petites les cellules de l'espace des phases, on n'arrivera jamais à un système dynamique discret qui puisse être décrit comme une permutation des cellules : la contraction de l'espace des phases entraine que certaines cellules ne seront jamais plus visitées même si on les a visitées au départ (par exemple parce que l'on a initié le mouvement à partir d'elles). Les mouvements se déroulent asymptotiquement sur un *attracteur* (qui est plus petit que tout l'espace des phases, bien que si on considère l'espace de phases comme continu l'attracteur pourrait être dense[19]).

Mais si on considère seulement les cellules sur lesquelles se déroule le mouvement on est dans une situation *identique à l'équilibre et hors équilibre*. On imagine que le mouvement est une permutation à un cycle, et donc il y aura un état stationnaire unique. Le temps pour parcourir le cycle sera, bien évidemment, toujours du même ordre de grandeur qu'à l'équilibre (dans des situations pas trop extrêmes des paramètres qui déterminent les forces agissantes sur le système) : donc la raison pour laquelle on peut espérer observer les moyennes temporelles et les calculer par intégration par rapport à une distribution de probabilité sur l'espace des phases reste la même que celle déjà discutée dans le cas d'équilibre (et liée à la loi des grands nombres).

[19] Ce qui montre seulement que la notion de "grandeur" d'un attracteur est plutôt délicate

Toutefois il y a une difficulté : c'est une difficulté qu'on aurait pu discuter déjà dans le cas de l'équilibre. On a supposé, sans critique, que les cellules de l'espace des phases étaient toutes égales. Mais même dans le cas de l'équilibre les systèmes sont chaotiques et donc toute cellule est déformée par l'évolution temporelle qui la dilate dans certaines directions et la contracte dans d'autres.

Il apparaît alors que la représentation du mouvement comme évolution d'une cellule vers une autre de forme et de taille identique est loin d'être triviale. Elle est en fait une hypothèse *forte* sur la dynamique, qui, à l'équilibre, sélectionne l'ensemble microcanonique comme distribution correcte à utiliser pour calculer les moyennes temporelles (et qui entraine le théorème de la chaleur). Il y a bien d'autres distributions invariantes sur l'espace des phases (contrairement à ce qu'on entend dire parfois) et l'hypothèse apparemment innocente que le mouvement se représente comme une permutation de cellules identiques en sélectionne une particulière.

Hors équilibre la difficulté devient plus manifeste. Car le volume des cellules ne reste même pas invariant contrairement au cas de l'équilibre (grâce au théorème de Liouville). De plus hors équilibre il faut s'attendre à ce que la représentation du mouvement comme évolution de cellules identiques conduise à sélectionner une distribution de probabilité particulière sur l'espace des phases, concentrée sur les cellules qui constituent l'attracteur, [Ga95a].

L'intérêt et l'importance des systèmes chaotiques au sens de l'hypothèse chaotique est que, en effet, pour tous ces systèmes *il y a une unique distribution stationnaire μ sur l'espace des phases qui donne les moyennes des grandeurs observées sur les mouvements qui commencent dans la grande majorité des cellules identiques en lesquelles on peut imaginer de diviser l'espace des phases.* C'est un résultat fondamental dû à Sinai et à Ruelle-Bowen : ainsi la distribution μ s'appelle *distribution SRB*, [Si68], [BR75]. Dans le cas de l'équilibre, elle coïncide avec la distribution microcanonique.

Ce n'est pas ici le lieu de poursuivre la critique de la vision discrète du mouvement, bien qu'elle soit intéressante ne fusse que pour une interprétation correcte des simulations numériques qui se font de plus en plus fréquentes, voir la note [20] page suivante.

L'hypothèse chaotique conduit naturellement à une représentation discrète différente du mouvement qui non seulement ne souffre pas des critiques qu'on vient de mentionner, mais qui nous donne une formule explicite pour la valeur des moyennes des observables, valable à la fois à l'équilibre (où elle se réduit à l'ensemble microcanonique) et hors équilibre.

Cette nouvelle représentation est aussi basée sur des cellules : mais elle ne sont pas vraiment petites dans le sens qu'elles sont considérablement plus grandes que les cellules que l'on a utilisées jusqu'à maintenant et qui avaient la taille minimale concevable. On peut donc les appeller "cellules à gros grains" ou grosses cellules, réservant le nom de cellules de taille fine aux précédentes.

Il est en effet possible de découper l'espace des phases en cellules

$E_1, E_2, \ldots = \{E_\kappa\}_{\kappa=1,\ldots}$ qui forment un *pavage* ou une *partition* $\mathcal{P}$ et qui ont la propriété de *covariance*.

Leur bords sont constitués par une réunion d'axes des systèmes locaux de coordonnées dont on a parlé plus haut : donc les bords consistent en des surfaces qui soit se contractent sous l'action de la dynamique, soit se dilatent. On dira que les frontières des cellules de la partition $\mathcal{P} = E_1, E_2, \ldots$ consistent en une partie qui se contracte ou "stable" et en une partie qui se dilate ou "instable". La propriété de covariance dit alors que sous l'action de l'évolution les cellules se déforment *mais les parties stables de leur bords évoluent de façon à terminer comme sous-ensembles de leur réunion* : la figure suivante illustre cette propriété simple.

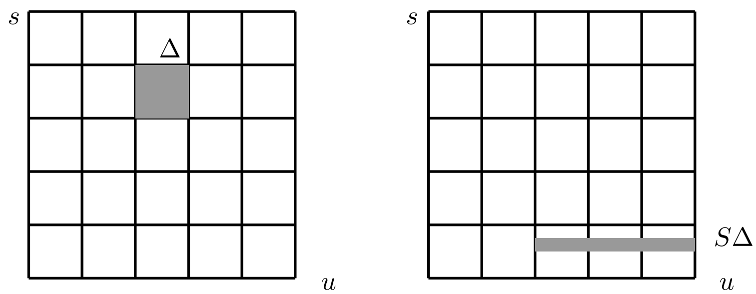


Fig. 9.A4.1

Si on a une telle partition (qui s'appelle *partition markovienne*) $\mathcal{P}$ on peut la raffiner en d'autres qui ont la même propriété de covariance : simplement en donnant un entier T et considérant la partition constituée par les ensembles $S^{-T}E_{\kappa_{-T}} \cap \ldots S^T E_{\kappa_T}$ qui, à cause de la contraction et de l'expansion de l'espace lors de l'évolution, forment une partition $\mathcal{P}_T$ dont les cellules deviennent aussi petites que l'on veut en prenant T assez grand.

Si F est une observable on peut en calculer la valeur moyenne simplement en considérant une partition markovienne $\mathcal{P}$ (arbitraire, car il n'y a pas d'unicité) en construisant la partition $\mathcal{P}_T$ avec T assez grand pour que F soit constant dans chaque cellule C de $\mathcal{P}_T$ et puis en posant :

9.*A*1.6
$$\langle F \rangle = \frac{\sum_C P(C)F(C)}{\sum_C P(C)} \tag{9.A1.6}$$

où $P(C)$ est un "poids" convenable. Il est construit en choisissant un point $c \in C$ et en considérant son évolution entre $-\tau$ et τ où τ est grand mais petit par rapport à T (par exemple $\tau = \frac{1}{2}T$).

On considère le point $S^{-\tau}c$ qui est transformé en $S^\tau c$ en un temps 2τ. On voit que l'axe, par $S^{-\tau}c$, des coordonnées qui se dilatent sous l'action de l'évolution est dilaté, au cours d'un temps 2τ, par un facteur qu'on appelle $\Lambda_{2\tau,i}(c)$: *alors le poids* $P(C)$ *peut être choisi égal à* $\Lambda_{2\tau,i}(c)^{-1}$.

L'équation (9.A1.6) est la formule qui remplace la distribution microcanonique hors équilibre : on peut prouver que l'on s'y ramène sous

l'hypothèse chaotique. La question qui se pose est si l'on peut tirer quelques conséquences générales de l'hypothèse chaotique moyennant l'usage de la représentation (9.A1.6) ci-dessus.[20]

Dans ce contexte, mentionnons que récemment on a réussi à déduire une conséquence qui apparemment à un certain intérêt. On va la formuler pour un système décrit par une équation différentielle (et donc en temps continu) $\dot{x} = f(x)$ qui engendre un flot $t \to S^t x$ dans l'espace des phases. On suppose aussi *que l'évolution est réversible* : c'est-à-dire qu'il y a une transformation isométrique I de l'espace des phases qui anti-commute avec l'évolution : $IS^t = S^{-t}I$.

Imaginons un système pour lequel l'hypothèse chaotique soit valable, donc décrit par une équation $\dot{x} = f(x)$ et soit $\sigma(x) = -\mathrm{div}\, f(x)$ la contraction de l'espace des phases associée. Supposons que l'on mesure la quantité $\sigma(S^n x)$ au cours du temps mais avec le système dans son état stationnaire. Appellons σ_+ sa moyenne temporelle *que l'on suppose non nulle* (alors elle

---

[20] L'expression (9.A1.6) pour la distribution SRB permet d'éclaircir la réprésentation de l'évolution comme permutation des cellules à taille fine. On doit imaginer que chaque élement C ("cellule à gros grain") de la partition markovienne $\mathcal{P}_T$, avec un T très grand de façon à ce que toute observable F (pertinente pour le comportment macroscopique) reste constante sur chaque C : pour une représentation fidèle du mouvement, on imagine que chaque C est quadrillé par des cellules très petites "de taille fine" *en nombre proportionnel à* $P(C)$. Par l'évolution les cellules de taille fine se répartissent entre les éléments C' de $\mathcal{P}_T$ qui intersectent SC. On fait évoluer de la même façon les autres cellules fines des éléments de $\mathcal{P}_T$: la théorie des distributions SRB montre que le nombre des cellules de taille fine qui viennent se trouver dans chaque $C \in \mathcal{P}_T$ ne change pas, à une très bonne aproximation près; c'est la stationnarité de la distribution SRB, [Ga95a]. Alors on peut définir l'évolution des cellules de taille fine simplement en disant qu'une cellule fine δ dans C évolue dans une des cellules fines qui sont dans la C' qui contient $S\delta$; il faut seulement faire attention à ne pas associer une même cellule fine de C' à deux cellules fines appartenant à differentes C (parmi celles telles que $SC \cap C' \neq \emptyset$) : on peut s'arranger de façon telle que la permutation des cellules fines ainsi définie soit à un seul cycle, car les détails du mouvement à l'intérieur des cellules C n'ont pas d'importance parce que les observables qui nous intéressent sont constantes dans les C. Mais la même construction peut être faite en remplaçant le poids $P(C)$ par $P(C)^\alpha$ avec $\alpha \neq 1$: on obtient ainsi d'autres distributions stationnaires *différentes* de la SRB, et on peut même en construire d'autres, [Si68], [Bo70]. On peut représenter de la même façon aussi ces autres distributions : mais on doit imaginer que les cellules de taille fine que l'on utilise pour en représenter une soient *différentes* de celles utilisées pour représenter les autres. *En fin de compte toutes les cellules fines ainsi introduites représentent l'attracteur.* Si on divise l'espace entier en (beaucoup de) cellules fines, de façon à ce que *toutes distributions stationnaires* puissent être représentées par une permutation des cellules fines qui se trouvent dans les $C \in \mathcal{P}_T$, alors on obtient une réprésentation discète très fidèle du mouvement. Mais toutes les cellules ne feront pas partie d'un cycle, car la dynamique est en général dissipative et une grande partie d'entre elles ne reviennent pas sur elles mêmes mais "tombent sur l'attracteur" où, dès lors, elles évoluent dans un cycle. La théorie de la distribution SRB montre que si on considère un ensemble ouvert dans l'espace des phases le comportement asymptotique du mouvement de tout point, sauf un ensemble de volume nul, est bien réprésenté par la distribution SRB, ce qui lui fait jouer un rôle particulier, au contraire des autres distributions que l'on peut definir : c'est-à-dire que la grande majorité (en volume) des cellules fines tombant sur l'attracteur vont se trouver parmi celles que l'on a associées aux cycles de la distribution SRB.

ne peut être que positive par un théorème général, [Ru96a]) et :

9.*A*1.7
$$p = \frac{1}{\tau\sigma_+}\int_{-\frac{1}{2}\tau}^{\frac{1}{2}\tau} \sigma(S^t x)dt \qquad (9.A1.7)$$

et soit $\pi_\tau(p) = e^{\tau\zeta(p)}$ la distribution de probabilité de cette observable. Alors :

9.*A*1.8
$$\frac{\zeta(p) - \zeta(-p)}{\tau\sigma_+} \equiv 1 \qquad (9.A1.8)$$

c'est le *théorème de fluctuation*, [GC95].

Je ne peux pas discuter ici la signification physique du théorème et de l'hypothèse de réversibilité, mais il est intéressant de souligner sa généralité, son indépendance du système considéré et aussi l'absence, dans sa formulation, de paramètres libres. Ce qui le rend en un certain sens analogue au théorème de la chaleur, qui lui aussi est général et sans paramètres libres.

Il suffira de dire que le théorème de fluctuation est une propriété qu'il faut quand même vérifier expérimentalement : en effet une partie de la théorie ci-dessus est née à la suite d'une expérience de simulation numérique et pour en interpréter théoriquement les résultats, [ECM93]. Il y a eu aussi quelques vérifications indépendantes, [BGG97], [LLP97].

La raison pour laquelle des expériences sont nécessaires est qu'il n'y a aucun espoir de prouver que des systèmes réels vérifient au sens mathématique du mot l'hypothèse chaotique; moins encore de prouver que des systèmes réel vérifient l'hypothèse ergodique. Il n'y a même pas d'espoir de prouver que des systèmes intéressants en simulation numérique ou dans la réalité vérifient des propriétés qui soient assez proches de celles des systèmes hyperboliques pour en déduire des conséquences telles que le théorème de fluctuation. Mais on peut croire que néanmoins "les choses se passent comme si l'hypothèse chaotique était littéralement vraie".

Il y a donc une nécessité d'un contrôle expérimental car on est dans la même situation qu'à l'équilibre : où tout en croyant, avec Feynman, que "*if we follow our solution* [*i.e.* motion] *for a long enough time it tries everything that it can do, so to speak*" (see p. 46-55 in [Fe63], vol. I), il a été néanmoins nécessaire de faire de bonnes vérifications expérimentales pour ne plus avoir de réserves ou de doutes sur l'hypothèse ergodique dans la théorie de l'équilibre.

Quelques références sont données ici pour guider le lecteur dans la littérature récente et ancienne. Elles sont loin d'être exhaustives : [HHP87], [EM90], [ECM90], [DPH96], [Ga95c], [Ga96a], [Ga97b], [Ge98], [GR97], [Ga98a], [Ga98b], [Ga96c], [Ga97a], [Ru97a], [Ru97c], [BG97], [MR97b], [Ku97].

Appendix 9.A2. Heuristic Derivation of the SRB Distribution

The discussion below follows [Ga95a],[Ga95d],[Ga98c], see also [Ga81]. A ball B containing a unit mass uniformly spread in it with density ρ and

centered around a fixed point[21] O for an Anosov map S which is a Anosov mixing map (hence such that the stable and unstable manifolds of O fill densely the phase space Ω on which S acts, see §9.4) will be elongated along the unstable manifold of O; in so doing the map S will compress the mass so that after T iterations the image $S^T B$ *will coat a large portion of* W_O^u *with a thin coating of mass.*

The mass around an infinitesimal surface element δ around a point $x \in W_O^u$ which is reached by the spreading coating, *i.e.* which has the form $x = S^T y$ for some y in the connected part of $W_O^u \cap B$ which contains O, will be the one that at time 0 was above the image $S^{-T}\delta$ of δ (of extremely small size, area $|\delta|$ and very close to the origin for T large and x fixed): *i.e.* it will be proportional to the area of the image times ρ and the proportionality constant will be essentially the power n of the radius h of the ball if n is the dimension of the stable manifold of O. In formulae the mass $d\mu$ in question will be:

9.A2.1
$$d\mu = \rho h^n \Lambda_{u,T}^{-1}(S^{-T}x)\,|\delta|\,. \tag{9.A2.1}$$

For T large we see therefore that the mass initially in B will coat a finite (but very large and increasing with T) surface of the (dense) manifold W_O^u. Hence we see that the SRB distribution, which should be the limit to which the described distribution of mass should tend, will be in a sense "concentrated" along the unstable manifold W_O^u.

Let δ, δ' be two (infinitesimal) surface elements on W_O^u centered around x, x' respectively with x, x' close and on the same stable manifold W_x^s. And suppose that the stable manifolds through the points of δ intersect δ' and vice-versa (*i.e.* δ, δ' are the bases of a "tube" whose generators are the stable manifolds). See Fig. 9.A5.1:

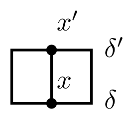


Fig. 9.A5.1

where the vertical lines represent stable manifolds and the horizontal parts of the unstable manifold. The surface elements δ, δ' are infinitesimal parts of a connected surface which, being very large and winding around on the manifold, "almost fills" the whole manifold. The surface (*i.e.* the unstable manifold of O) is not drawn; it would connect the two surface elements. We see that the ratio of the masses originally in B and coating δ and δ' is

9.A2.2
$$\frac{\Lambda_{u,T}^{-1}(S^{-T}x)}{\Lambda_{u,T}^{-1}(S^{-T}x')}\frac{|\delta|}{|\delta'|} \tag{9.A2.2}$$

where $|\delta|, |\delta'|$ denote the *surface areas* of the elements δ, δ'; but the ratio

[21] It is not restrictive to suppose that there is a fixed point for S. In fact Anosov systems always admit periodic points: they are always dense on phase space $\mathcal{F}$. If ω is a periodic point of period N then it is a fixed point for S^N. Clearly the map S^N is still an Anosov map, so that the following discussion will apply to the map S^N and even to S provided we take T an integer multiple of N.

$|\delta|/|\delta'|$ can be expressed as

9.*A*2.3
$$\frac{|\delta|}{|\delta'|}=\frac{S^{-T}(S^T\delta)}{S^{-T}(S^T\delta')}=\frac{\Lambda^{-1}_{u,T}(S^Tx)}{\Lambda^{-1}_{u,T}(S^Tx')}\frac{|S^T\delta|}{|S^T\delta'|} \qquad (9.A2.3)$$

and by the composition rule of derivatives (and Jacobian determinants) we see that the (9.A2.2) can be written

9.*A*2.4
$$\frac{\Lambda^{-1}_{u,T}(S^{-T}x)\Lambda^{-1}_{u,T}(S^Tx)}{\Lambda^{-1}_{u,T}(S^{-T}x')\Lambda^{-1}_{u,T}(S^Tx')}\frac{|S^T\delta|}{|S^T\delta'|} \qquad (9.A2.4)$$

but the last ratio approaches 1 as $T\to\infty$ because δ,δ' are infinitesimal and $S^T\delta, S^T\delta'$ get closer and closer as $T\to\infty$. The limit as $T\to\infty$ is

9.*A*2.5
$$\lim_{T\to\infty}(9.A2.2) = \prod_{k=-\infty}^{\infty}\frac{\Lambda^{-1}_{1,u}(S^{-k}x)}{\Lambda^{-1}_{1,u}(S^{-k}x')}\,. \qquad (9.A2.5)$$

It is not difficult to check that the infinite product on the right converges because the points S^kx, S^kx' tend to get close exponentially fast *both* in the future and in the past, being at the same time on the stable and on the unstable manifolds of each other.

Thus the coating generated by the splashing (due to the time evolution) of the mass initially in the ball B of the unstable manifold W^u_O will have the *formal* density $\prod_{k=-\infty}^{\infty}\Lambda^{-1}_{u,1}(S^{-k}x)$ which can be written (formally) as e^{-H} with $H=\sum_{k=-\infty}^{\infty}\log\Lambda_{u,1}(S^{-k}x)$. Since x can be represented as a sequence $\underline{\sigma}$ via its symbolic coding, $x=x(\underline{\sigma})$, on a Markov partition $\mathcal{E}$ we can define $\lambda_u(\underline{\sigma})=\log\Lambda_{u,1}(x(\underline{\sigma}))$ and obtain that the SRB distribution μ will be represented as a distribution on the space of the (compatible) sequences associated with the Markov partition and it will have the formal expression

9.*A*2.6
$$\mu(d\underline{\sigma}) = \mathit{const}\, e^{-\sum_{k=-\infty}^{\infty}\lambda_u(\vartheta^k\underline{\sigma})} \qquad (9.A2.6)$$

if ϑ is the shift on the bilateral sequences $\underline{\sigma}$.

Recalling §5.10 we see that the distribution μ can be interpreted as a Gibbs state for a one-dimensional Ising model with short-range interaction in the sense of the discussion in Chap.V following (5.10.12), see §5.10,.

The function $\lambda_u(\underline{\sigma})$ has in fact all the properties needed for the latter interpretation, as a consequence of the discussion in §9.5. The above heuristic discussion establishes the connection between the statistical mechanics of one-dimensional spin systems with short-range interactions and the apparently highly nontrivial dynamics of an Anosov system. The above remarks lay the foundations of the thermodynamic formalism, [Si68],[Ru76],[Bo74],[Ru78b].

In a sense the reduction of the system to an Ising model is the chaotic dynamics analogue of the *integration by quadratures* of classical mechanics.

Appendix 9.A3. Aperiodic Motions Can be Begarded as Periodic with Infinite Period!

This famous and criticized statement of Boltzmann, [Bo66], is the heart of the application of the heat theorem for monocyclic systems to a gas in a box. Imagine the box containing the gas to be covered by a piston of section A and located to the right of the origin at distance L, so that $V = AL$.

The microscopic model for the piston will be a potential $\overline{\varphi}(L-\xi)$ if $x = (\xi,\eta,\zeta)$ are the coordinates of a particle. The function $\overline{\varphi}(r)$ will vanish for $r > r_0$, for some $r_0 < L$, and diverge to $+\infty$ at $r = 0$. Thus r_0 is the width of the layer near the piston where the force of the wall is felt by the particles that happen to roam there.

Noting that the potential energy due to the walls is $\varphi = \sum_j \overline{\varphi}(L-\xi_j)$ and that $\partial_V\varphi = A^{-1}\partial_L\varphi$ we must evaluate the time average of

9.A3.1
$$\partial_L\varphi(x) = -\sum_j \overline{\varphi}'(L-\xi_j)\,. \tag{9.A3.1}$$

As time evolves the particles with ξ_j in the layer within r_0 of the wall will feel the force exercised by the wall and bounce back. Fixing the attention on one particle in the layer we see that it will contribute to the average of $\partial_L\varphi(x)$ the amount

9.A3.2
$$\frac{1}{\text{total time}}2\int_{t_0}^{t_1} -\overline{\varphi}'(L-\xi_j)dt \tag{9.A3.2}$$

if t_0 is the first instant when the point j enters the layer and t_1 is the instant when the ξ-component of the velocity vanishes "against the wall". Since $-\overline{\varphi}'(L-\xi_j)$ is the ξ-component of the force, the integral is $2m|\dot\xi_j|$ (by Newton's law), provided $\dot\xi_j > 0$ of course. One assumes that the density is low enough so that no collisions between particles occur while the particles travel within the range of the potential of the wall: *i.e.* the mean free path is much greater than the range of the potential $\overline{\varphi}$ defining the wall.

The number of such contributions to the average per unit time is therefore given by $\rho_{wall}\,A\int_{v>0}2mv\,f(v)\,v\,dv$ if ρ_{wall} is the density (average) of the gas near the wall and $f(v)$ is the fraction of particles with velocity between v and $v+dv$. Using the ergodic hypothesis (*i.e.* the microcanonical ensemble) and the equivalence of the ensembles to evaluate $f(v)$ it follows that:

9.A3.3
$$p \stackrel{def}{=} -\langle\partial_V\varphi\rangle = \rho_{wall}\beta^{-1} \tag{9.A3.3}$$

where $\beta^{-1} = k_BT$ with T the absolute temperature and k_B Boltmann's constant. Hence we see that (9.A3.3) yields the correct value of the pressure, see Chap.I, Chap.II; in fact it is often even taken as the microscopic definition of the pressure, [MP72].

On the other hand we have seen in §9.1, (9.1.7) (repeating the analysis in Appendix 1.A1, Chap.I), that if all motions are periodic the quantity p

in (9.A3.3) is the right quantity that would make the heat theorem work. Hence regarding all trajectories as periodic (*i.e.* the system as monocyclic) leads to the heat theorem with p, U, V, T having the *right physical interpretation*. And Boltzmann thought since the beginning of his work that trajectories confined into a finite region of phase space could be regarded as periodic possibly with infinite period, [Bo66].

Appendix 9.A4. Gauss' Least Constraint Principle

Let $\varphi(\underline{\dot x},\underline{x})=0$, $(\underline{\dot x},\underline{x})=\{\underline{\dot x}_j,\underline{x}_j\}$ be a constraint and let $\underline{R}(\underline{\dot x},\underline{x})$ be the constraint reaction and $\underline{F}(\underline{\dot x},\underline{x})$ the active force.

Consider all the possible accelerations $\underline{a}$ compatible with the constraints and a given initial state $\underline{\dot x},\underline{x}$. Then $\underline{R}$ is *ideal* or *satisfies the principle of minimal constraint* if the actual accelerations $\underline{a}_i=\frac{1}{m_i}(\underline{F}_i+\underline{R}_i)$ minimize the *effort*

9.A4.1
$$\sum_{i=1}^N\frac{1}{m_i}(\underline{F}_i-m_i\underline{a}_i)^2\longleftrightarrow\sum_{i=1}^N(\underline{F}_i-m_i\underline{a}_i)\cdot\delta\underline{a}_i=0 \qquad (9.A4.1)$$

for all possible variations $\delta\underline{a}_i$ compatible with the constraint φ. Since all possible accelerations following $\underline{\dot x},\underline{x}$ are such that $\sum_{i=1}^N\partial_{\underline{\dot x}_i}\varphi(\underline{\dot x},\underline{x})\cdot\delta\underline{a}_i=0$ we can write

9.A4.2
$$\underline{F}_i-m_i\underline{a}_i-\alpha\,\partial_{\underline{\dot x}_i}\varphi(\underline{\dot x},\underline{x})=\underline{0} \qquad (9.A4.2)$$

with α such that

9.A4.3
$$\frac{d}{dt}\varphi(\underline{\dot x},\underline{x})=0, \qquad (9.A4.3)$$

i.e.

9.A4.4
$$\alpha=\frac{\sum_i(\underline{\dot x}_i\cdot\partial_{\underline{x}_i}\varphi+\frac{1}{m_i}\underline{F}_i\cdot\partial_{\underline{\dot x}_i}\varphi)}{\sum_i m_i^{-1}(\partial_{\underline{\dot x}_i}\varphi)^2} \qquad (9.A4.4)$$

which is the analytic expression of the Gauss' principle, see [Wi89].

Note that if the constraint is even in the $\underline{\dot x}_i$ then α is odd in the velocities: therefore if the constraint is imposed on a system with Hamiltonian $H=K+V$, with K quadratic in the velocities and V depending only on the positions, and if other purely positional forces (conservative or not) act on the system then the resulting equations of motion are reversible if time reversal is simply defined as velocity reversal.

The Gauss' principle has been somewhat overlooked in the Physics literature in statistical mechanics: its importance has again only recently been brought to the attention of researchers, see the review [HHP87]. A notable, though by now ancient, exception is a paper of Gibbs, [Gi81], which develops variational formulas which he relates to Gauss' principle of least constraint.

Conceptually this principle should be regarded as a *definition* of *ideal nonholonomic constraint*, much as D'Alembert's priciple or the least action principle are regarded as the definition of *ideal holonomic constraint*.

Bibliography

Strict lexicographic ascii order: **first digits, second capitals, third lower cases.** Coauthored papers or papers whose authors have a composite name are marked by the capital initials of the authors or of the names; single author papers or are marked by the capital initial followed by the first following letter in lower case.

[AA68] Arnold, V.I., Avez, A.: *Ergodic problems of classical mechanics*, Benjamin, New York, 1968.

[ACG96] Artuso, R., Casati, G., Guarneri, I., *Numerical Experiments on Billiards*, Journal of Statistical Physics, **83**, 145–166, 1996.

[AGM71] Abraham, D, Gallavotti, G., Martin–Löf, A.: *Surface tension in the Ising model*, Lettere al Nuovo Cimento, **2**, 143–146, 1971. And *Surface tension in the two–dimensional Ising model*, Physica, 65, 73–88, 1973.

[AL95] Abraham, D.B., Latrémolière, F.T.: *Corner spontanaous magnetization*, Journal of Statistical Physics, **81**, 539–559, 1995.

[AM73] Abraham, D.B., Martin-Löf, A.: *The transfer matrix for a pure phase in the two dimensional Ising model*, Communications in Mathematical Physics, **31**, 245–268, 1973.

[AR76] Abraham, D.B., Reed, P: *Interface profile in the two dimensional Ising model*, Communications in Mathematical Physics, **49**, 35–46, 1976.

[AW68] Adler, R., Weiss, B.: *Similarity of automorphisms of the torus*, Memoires of the American Mathematical Society, **98**, 1968.

[Ab71] Abraham, D.B.: *On the transfer matrix for the two dimensional Ising model*, Studies in Applied Mathematics, **50**, 71–88, 1971.

[Ab78a] Abraham, D.B.: *n–point functions for the rectangular Ising ferromagnet*, Communications in Mathematical Physics, **60**, 205–213, 1978.

[Ab78b] Abraham, D.W.: *Block spins in the edge of an Ising ferromagnet in a half plane*, Journal of Statistical Physics, **19**, 553–556, 1978.

[Ab84] Abraham, D.W.: *intrinsic structure and long ranged correlations in interfaces*, Physical Review, **29B**, 526–526, 1984.

[Ai80] Aizenman, M.: *Translation invariance and instability of phase coexistence in the two dimensional Ising system*, Communications in Mathematical Physics, **73**, 83–94, 1980.

[An82] Andrej, L.: *The rate of entropy change in non–Hamiltonian systems*, Physics Letters, **111A**, 45–46, 1982. And *Ideal gas in empty space*, Nuovo Cimento, **B69**, 136–144, 1982. See also *The relation between entropy production and K–entropy*, Progress in Theoretical Physics, **75**, 1258–1260, 1986.

[An84] Anderson, P.W.: *Basic notions of condensed matter physics*, Benjamin–Cummings, Menlo Park, California, 1984.

[As70] Asano, T.: *Theorems on the partition function of the Heisenberg ferromagnet*, Journal Physical Society Japan, **29**, 350–359, 1970.

[BC94] Borgs, C., Chayes, J.: *The covariance matrix in the Potts model: a random cluster analysis*, Journal of Statistical Physics, **82**, 1235–1297, 1995.

[BCS57] Bardeen, J., Cooper, L.N., Schrieffer, J.R.: *Theory of superconductivity*, Physical Review, **108**, 1175–1204, 1957.

[BF67] Burford, R.J., Fisher, M.E.: *Theory of critical scattering and correlations. I. The Ising model*, Physical Review, **156**, 583-622, 1967.

[BG95] Benfatto, G., Gallavotti, G.: *Renormalization group*, Princeton University Press, 1995.

[BG97] Bonetto, F., Gallavotti, G.: *Reversibility, coarse graining and the chaoticity principle*, Communications in Mathematical Physics, **189**, 263–276, 1997.

[BGG84] Benettin, G., Galgani, L., Giorgilli, A.: *Boltzmann's ultraviolet cut-off and Nekhoroshev's theorem on Arnold diffusion*, Nature, **311**, 444-445, 1984.

[BGG93] Benettin, G., Galgani, L., Giorgilli, A.: *The dynamical foundations of classical statistical mechanics and the Boltzmann-Jeans conjecture*, Edited by S. Kuksin, V.F. Lazutkin, J. Pöschel, Birkhauser, 1993.

[BGG97] Bonetto, F., Gallavotti, G., Garrido, P.: *Chaotic principle: an experimental test*, Physica D, **105**, 226–252, 1997.

[BGJS73] Benettin, G., Gallavotti, G., Jona–Lasinio, G., Stella, A.: *On the Onsager–Yang value of the spontaneous magnetization*, Communications in Mathematical Physics: **30**, 45-54, 1973,

[BH93] Böhm, D., Hiley, B.J.: *The undivided universe: an ontological interpretation of Quantum Mechanics*, Routledge & Keegan, London, 1993.

[BJS72] Benettin, G., Jona–Lasinio, G., Stella, A.: *Duality and Correlation Functions in the Ising Model*, Lettere Nuovo Cimento, **4**, 443-447, 1972.

[BK52] Berlin, T.H., Kac, M.: *The spherical model of a ferromagnet*, Physical Review, **86**, 821–835, 1952.

[BLPO79] Bricmont, J., Lebowitz, J.L., Pfister, C., Olivieri, E.: *Nontranslation-invariant Gibbs states with coexisting phases*, Communications in Mathematica Physics, **66**, 1–20, 1979; and the first three authors *Nontranslation-invariant Gibbs states with coexisting phases*, Communications in Mathematical Physics, **66**, 21–36, 1979 and *Absence of symmetry breakdown and uniqueness of the vacuum for multicomponent field theories*, Communications in Mathematical Physics, **69**, 267–291, 1979.

[BR75] Bowen, R., Ruelle, D.: *The ergodic theory of axiom A flows*, Inventiones Mathematicae, **29**, 181–205, 1975.

[BS67] Berezin, F.A., Sinai, Y.G.: *The existence of phase transition for a lattice gas with attraction between the particles*, Transactions of the Moscow Mathematical Society, **19**, 219–234, 1967.

[Ba00] Bachelier,, L.: *Théorie de la spéculation*, Annales Scientifiques de l' École Normale Superieure, 17, 21-36, 1900.

[Ba82] Baxter, R.: *Exactly solved models*, Academic Press, London, 1982.

[Ba90] Bach, A.: *Boltzmann's probability distribution of 1877*, Archive for the History of exact sciences, **41**, 1-40, 1990.

[Be31] Bethe, H.A.: *Zür theorie der Metalle. I. Eigenwerte und Eigenfunktionen der linearen Atomkette*, Zeitschrift für Physik, **71**, 205–226, 1931.

[Be69] Berezin, F.A.: *The plain Ising model*, Russian Mathematical Surveys, **24**, 2, 1-22, 1969.

[Be87] Bell, J.S.: *Speakable and unspeakable in quantum mechanics*, Cambridge U. Press, Cambridge, 1987.

[Be84] Benettin, G.: *On the Boltzmann conjecture for a classical "freezing" of fast degrees of freedom*, in "Boltzmann's legacy 150 years after his birth, Atti dei Convegni Lincei, **131**, Accademia dei Lincei, 89-105, 1997.

[Be94] Benettin, G.: *Ergodicity and time-scales for statistical equilbrium in classical dynamical systems*, Progress of Theoretical Physics, Supplement, **116**, 207-234, 1994.

[Bo64] Boltzmann, L.: *Lectures on gas theory*, English edition annotated by S. Brush, University of California Press, Berkeley, 1964, (reprint).

[Bo66] Boltzmann, L.: *Über die mechanische Bedeutung des zweiten Haupsatzes der Wärmetheorie*, in "Wissenschaftliche Abhandlungen", ed. F. Hasenöhrl, vol. I, p. 9–33, reprinted by Chelsea, New York.

[Bo68] Boltzmann, L.: *Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten*, in "Wissenschaftliche Abhandlungen", ed. F. Hasenöhrl, vol. I, p. 49–96, reprinted by Chelsea, New York.

[Bo70] Bowen, R.: *Markov partitions for Axiom A diffeomorphisms*, American Journal of Mathematics, **92**, 725–747, 1970. And: *Markov partitions and minimal sets for Axiom A diffeomorphisms*, American Journal of Mathematics, **92**, 907–918, 1970.

[Bo71a] Boltzmann, L.: *Analytischer Beweis des zweiten Hauptsatzes der mechanischen Wärmetheorie aus den Sätzen über das Gleichgewicht des lebendigen Kraft*, in "Wissenschaftliche Abhandlungen", ed. F. Hasenöhrl, vol. I, p. 288–308, reprinted by Chelsea, New York.

[Bo71b] Boltzmann, L.: *Einige allgemeine Sätze über Wärmegleichgewicht*, in "Wissenschaftliche Abhandlungen", ed. F. Hasenöhrl, vol. I, p. 259–287, reprinted by Chelsea, New York.

[Bo71c] Boltzmann, L.: *Zusammenhang zwischen den Sätzen über das Verhalten mehratomiger Gasmoleküle mit Jacobi's Princip des letzten Multiplicators*, this is sometimes quoted as a paper but it is sec. 1 of [Bo71b], see the introduction of [Gi81].

[Bo72] Boltzmann, L.: *Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen*, english translation in S. Brush, *Kinetic theory*, Vol. 2, p. 88. Original in "Wissenschaftliche Abhandlungen", ed. F. Hasenöhrl, vol. I, p. 316–402, reprinted by Chelsea, New York.

[Bo74] Boltzmann, L.: *Theoretical Physics and philosophical writings*, ed. B. Mc Guinness, Reidel, Dordrecht, 1974.

[Bo77] Boltzmann, L.: *Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respektive den Sätzen über das Wärmegleichgewicht*, in "Wissenschaftliche Abhandlungen", vol. II, p. 164–223, F. Hasenöhrl, Chelsea, New York, 1968 (reprint).

[Bo84] Boltzmann, L.: *Über die Eigenshaften monzyklischer und anderer damit verwandter Systeme*, in "Wissenshafltliche Abhandlungen", ed. F.P. Hasenöhrl, vol. III, p. 122–152, Chelsea, New York, 1968, (reprint).

[Bo96] Boltzmann, L.: *Entgegnung auf die wärmetheoretischen Betrachtungen des Hrn. E. Zermelo*, engl. transl.: S. Brush, "Kinetic Theory", vol. 2, 218, Pergamon Press, Oxford, 1965-1972.

[Bo97] Boltzmann, L.: *Abhandlung Über die mechanische Erklärung irreversibler Vorgänge*, engl. trans. in S. Brush, "Kinetic Theory", vol. 2, 238, Pergamon Press, Oxford, 1965-1972.

[Bo97b] Bonetto, F.: private communication.

[Br65] Brout, R.H.: *Phase transitions*, Benjamin, New York, 1965.

[Br66] Brush, S.G.: *Kinetic theory*, Pergamon, Oxford, 1966.

[Br69] Brush, S.G.: *For a history of the Ising model*, Reviews of Modern Physics, **39**, 883 (1969).

[Br76] Brush, S.: *The kind of motion we call heat*, North Holland, Amsterdam, 1976 (vol. II), 1986 (vol. I).

[Br86] Brydges, D.,C,: *A short course on cluster expansions*, in "Critical phenomena, random systems, gauge theories", ed. K. Osterwalder, R. Stora, Les Houches XLIII, North Holland, 1986.

[CC53] Chapman S., Cowling T.: *The mathematical Theory of Nonuniform Gases*, Cambridge, 1953.

[CCO74] Capocaccia, D., Cassandro, M., Olivieri, E.: *A study of metastability in the Ising model*, Communications in Mathematical Physics, **39**, 185-205, 1974.

[CELS93] Chernov, N. I., Eyink, G. L., Lebowitz, J.L., Sinai, Y.: *Steady state electric conductivity in the periodic Lorentz gas*, Communications in Mathematical Physics, **154**, 569–601, 1993.

[CF71] Camp, W.J., Fisher, M.E.: *Behavior of two point correlation functions at high temperature*, Physical Review Letters, **26**, 73–77 and 565–568, 1971.

[CL98] Ciliberto, S., Laroche, C.: *An experimental verification of the Gallavotti–Cohen fluctuation theorem*, Journal de Physique, **8**, 215–222, 1998.

[Ca67] Cantoni, G.: *Su alcune condizioni fisiche dell' affinità e sul moto browniano*, Il Nuovo Cimento, **27**, 156–167, 1867.

[Ca83] Cammarota, C.: *Decay of correlations for infinite range interactions in unbounded spin systems*, Communications in Mathematical Physics, **85**, 517-528, 1983.

[Ce69] Cercignani, C.: *Mathematical methods in kinetic theory*, Plenum, New York, 1969.

[Cl65] Clausius, R.: *The nature of the motion which we call heat*, translated in *Kinetic Theory*, ed. S. Brush, Vol. 1, p. 111—134, Pergamon, Oxford, 1965.

[Co69] Cohen, E.G.D.: *The kinetic theory of dilute gases*, in "Transport Phenomena in Fluids", H. Hanley ed., Ch. VI, 119–156; Dekker, New York, 1969.

[Co93] Cohen, E.G.D.: *Fifty years of kinetic theory*, Physica A, **194**, 229–257, 1993.

[DKS92] Dobrushin, R.L., Kotecky, R., Shlosman, S.: *The Wulff construction*, Tnaslations of Mathematical Monograohs, **104**, American Mathematical Society, Providence, 1992. See also: Dobrushin, R.L., Kotecky, R., Shlosman, S.: *A microscopi justification of the Wulff construction*, Journal of Statistical Physics, **72**, 1–14, 1993; and Dobrushin, R.L., Hryniv, O.: *Fluctuations of the phase boundary in the 2D Ising model*, Communications in Mathematical Physics, **189**, 395–445, 1997.

[DL67] Dyson, F., Lenard, A.: *Stability of matter, I*, Journal of Mathematical Physics, **8**, 423–434, 1967. And *Stability of matter, II*, Journal of Mathematical Physics, **9**, 698–711, 1968.

[DLS78] Dyson, F.J., Lieb, E.H., Simon, B.: *Phase transitions in quantum spin systems with isotropic and anisotropic interactions*, Journal of Statistical Physics, **18**, 335–383, 1978.

[DM96] Dettman, C.P., Morris, G.P.: *Proof of conjugate pairing for an isokinetic thermostat*, Physical Review **53 E**, 5545–5549, 1996.

[DM67] Dobrushin, R., Minlos, R.: *Existence and continuity of the pressure in classical Statistical Mechanics*, Theory of probability and its applications, **12**, 535–559, 1967.

[DPH96] Dellago, C., Posch, H., Hoover, W.: *Lyapunov instability in system of hard disks in equilibrium and non-equilibrium steady states*, Physical Review, **53E**, 1485–1501, 1996.

[Do68a] R.L. Dobrushin, R.L.: *The description of a random field by means of conditional probabilities and conditions of its regularity*, Theory of probability and its applications, **13**, 197–224, 1968.

[Do68b] R.L. Dobrushin, R.L.: *Gibbsian random fields for lattice systems with pairwise interactions*, Functional Analysis and Applications, **2**, 31–43, 1968; and *Problem of uniqueness of a Gibbs random field and phase transition*, Functional Analysis and Applications, **2**, 44–57, 1968;

[Do68c] Dobrushin, R.L.: *Gibbs field: the general case*, Functional Analysis and Applications, **3**, 27–35, 1969.

[Do72] Dobrushin, R.L.: *Gibbs state describing coexistence of phases for a three dimensional Ising model*, Theory of probability and its applications, **17**, 582–600, 1972. And *Investigation of Gibbsian states for three–dimensional lattice systems*, Theory of probability and its applications, **18**, 253–271, 1973. Both reprinted in [Si91].

[Do80] Dobrushin, R.L.: *Automodel generated random fields and their renormalization group*, in *Multicomponent random fields*, Advances in Probability and related topics, **6**, 153–198, 1980.

[Do98] Dorfman, J.R.: *An introduction to chaos in non-equilibrium statistical mechanics*, p. 1–287, University of Maryland, preprint, 1998.

[Dr88] Dressler, U.: *Symmetry property of the Lyapunov exponents of a class of dissipative dynamical systems with viscous damping*, Physical Review, **38A**, 2103–2109, 1988.

[Du59] Dugas, R.: *La théorie phisique au sens de Boltzmann*, Griffon, Neuchâtel, 1959.

[Dy69] Dyson, F.: *Existence of a phase transition in a one–dimensional Ising ferromagnet*, Communications in Mathematical Physics, **12**, 91–107, 1969.

[ECM90] Evans, D.J.,Cohen, E.G.D., Morriss, G.P.: *Viscosity of a simple fluid from its maximal Lyapunov exponents*, Physical Review, **42A**, 5990–5997, 1990.

[ECM93] Evans, D.J.,Cohen, E.G.D., Morriss, G.P.: *Probability of second law violations in shearing steady flows*, Physical Review Letters, **71**, 2401–2404, 1993.

[EE11] Ehrenfest, P., Ehrenfest, T.: *The conceptual foundations of the statistical approach in Mechanics*, Dover, New York, 1990, (reprint).

[EG75] Esposito, R., Gallavotti, G.: *Approximate simmetries and their spontaneous breakdown*, Annales de l'Institut Henri Poincarè: **22**, 159–172, 1975.

[EGM98] Esposito, R., Gallavotti, G., Marra, R.: *Boltzmann's equation and thermostats*, preprint, Roma, 1998.

[EM90] Evans, D.J., Morriss, G.P.: *Statistical Mechanics of Nonequilibrium fluids*, Academic Press, New York, 1990.

[ER85] Eckmann, J.P., Ruelle, D.: *Ergodic theory of chaos and strange attractors*, Reviews of Modern Physics, **57**, 617–656, 1985.

[ES93] Evans, D.J., Sarman, S.: *Equivalence of thermostatted nonlinear response*, Physical Review E, **48 E**, 65–70, 1993.

[ES94] Evans, D., Searles, D.: *Equilibrium microstates which generate the second law violating steady states*, Physical Review E, **50E**, 1645–1648, 1994.

[Ei56] Einstein, A.: *Investigations on the theory of the Brownian movement*, edited by R. Fürth, A. Cowper, Dover, New York, 1956.

[Ev93] Evans, D.J.: *The equivalence of Norton and Thévenin ensembles*, Molecular Physics, **80**, 221–224, 1993.

[FKG71] Fortuin, C.M., Kasteleyn, P.W., Ginibre, J.: *Correlation inequalities on some partially ordered sets*, Communications in Mathematical Physics, **22**, 89–104, 1971.

[FPU55] Fermi, E., Pasta, J., Ulam, S.: *Studies of nonlinear problems*, Los Alamos report LA-1940, 1955, printed in E., Fermi, *Collected papers*, p. 978–988, vol. **2**, 1965.

[FR66] Fisher, M.E., Ruelle, D.: *The stability of many particles systems*, Journal of Mathematical Physics, **7**, 260–270, 1966.

[FS62] Fisher, M.E., Sykes, M.F.: *Antiferromagnetic susceptibility of the simple cubic and body centered cubic Ising lattice*, Physica, **28**, 939–956, 1962.

[Fe63] Feynman, R.P., Leighton, R.B., Sands, M.: *Lectures in Physics.*, Vol. I,II,III, Addison Wesley, Reading, Massachussets, 1963.

[Fe72] Feynman, R.P.: *Statistical mechanics*, Benjamin, Reading, Massachussets, 1972.

[Fe85] Fefferman, C.: *The atomic and molecular nature of matter*, Revista Matemática Iberoamericana, **1**, 1–44, 1985.

[Fi64] Fisher, M.E.: *The free energy of a macroscopic system*, Archive for Rational Mechanics and Analysis, **17**, 377–410, 1964.

[Fi65] Fisher, M. E.: *Lectures in Physics*, vol. **7C** Boulder, Colorado, 1965.

[Fi67a] Fisher M.E.: *Theory of equilibrium critical phenomena*, Reports on Progress in Theoretical Physics, **30**, 615, 1967.

[Fi67b] Fisher, M. E.: , Physical Review, **162**, 475 (1967).

[Fi67c] Fisher, M. E.: *Theory of condensation and the critical point*, Physics–Physica–Fyzika, **3**, 255–283, 1967.

[Fi98] Fisher, M.E.: *Renormalization group theory: its basis and formulation in statistical physics*, Reviews of Modern Physics, 653–681, 1998.

[Fr81] Frölich, J.: *On the triviality of $\lambda\phi_d^4$ theories and the approach to the critical point in ($d \geq 4$) dimensions*, Nuclear Physics, **200 B**, 281–296, 1981.

[Fr86] Frölich, J.: *Mathematical aspects of the Physics of disordered systems*, Critical phenomena, random systems, gauge theories, ed. K. Osterwalder, R. Stora, p. 725–893, North Holland, Amsterdam, 1986.

[Fr97] Frisch, U.: Turbulence, Cambridge U. Press, 1997.

[GC95] Gallavotti, G., Cohen, E.G.D.: *Dynamical ensembles in non-equilibrium statistical mechanics*, Physical Review Letters, **74**, 2694–2697, 1995. Gallavotti, G., Cohen, E.G.D.: *Dynamical ensembles in stationary states*, Journal of Statistical Physics, **80**, 931–970, 1995.

[GG94] Gallavotti, G., Garrido, P.: *Billiards correlation functions*, Journal of Statistical Physics, **76**, 549–586, 1994.

[GH64] Green, H.S., Hurst, C.H.: Order Disorder Phenomena, Wiley, New York, 1964.

[GHM77] Gruber, C., Hintermann, A., Messager, A., Miracle-Sole, S.: *Uniqueness of the invariant equilibium state and surface tension*, Communications in Mathematical Physics **56**, 147–159, 1977.

[GJ75] Gallavotti, G., Jona, G.: *Limit theorems for multidimensional Markov processes*, Communications in Mathematical Physics: **41**, 301–307, 1975.

[GL71] Gallavotti, G., Lebowitz, J.L.: *Phase transitions in binary lattice gases* , Journal of Mathematical Physics, **12**, 1129–1133, 1971.

[GLL70] G. Gallavotti, O. Lanford, J. Lebowitz: *Thermodynamic limit of time–dependent correlation functions for one–dimensional systems*: Journal of Mathematical Physics: **11**, 2898–2905, 1970.

[GM72a] Gallavotti, G., Miracle–Solé, S. *Equilibrium states of the Ising Model in the Two-phases Region*, Physical Review, **5**B, 2555–2559, 1972.

[GM72b] Gallavotti, G., Martin–Löf A.: *Surface tension in the Ising Model*, Communications in Mathematical Physics: **25**, 87–126, 1972.

[GMM72] Gallavotti, G., Martin-Löf, A., Miracle–Solé, S: *Some problems connected with the description of coexisting phases at low temperatures in the Ising model*, in Springer lecture notes in Physics, edited by A. Lenard, vol. **20**, p. 159–204, Berlin 1972.

[GMR69] Gallavotti, G., Miracle–Solé, S., Ruelle, D.: *Absence of phase transitions in one dimensional systems with hard core*, Physics letters, **26A**, 350–351, 1968. And Gallavotti, G., Miracle–Solé, S.: *Absence of phase transitions in hard core one–dimensional systems with long range interactions*, Journal of Mathematical Physics: **11**, 147–154, 1970.

[GR97] Gallavotti, G., Ruelle, D.: *SRB states and non-equilibrium statistical mechanics close to equilibrium*, Communications in Mathematical Physics, **190**, 279–285, 1997.

[GS72] Galgani, L., Scotti, A.: *Planck-like distributions in classical nonlinear mechanics*, Physical Review Letters, **28**, 1173-1176, 1972.

[GV72] Gallavotti, G., Van Beijeren, H.: *The Phase separation line in the two— dimensional Ising model and its relationships with the non–free random walk problem*, Lettere al Nuovo Cimento, **4**, 699–703, 1972.

[Ga69] G. Gallavotti, *Time evolution problems in Classical Statistical Mechanics and the Wind–tree–model*: in "Cargese Lectures in Physics", vol. IV, ed. D. Kastler, Gordon Breach, Paris, 1970, pp.257–275; see p. 271–272, formula (6.2). And G. Gallavotti, *Divergences and approach to equilibrium in the Lorentz and the Wind–tree models*, Physical Review, **185**, 308–322, 1969.

[Ga72a] Gallavotti, G.: *Phase separation line in the two–dimensional Ising model*: Communications in Mathematical Physics, **27**, 103–136, 1972.

[Ga72b] Gallavotti, G.: *Instabilities and phase transitions in the Ising model*, La Rivista del Nuovo Cimento, **2**, 133–169, 1972.

[Ga76] Gallavotti, G.: *Probabilistic aspects of critical fluctuations*, in Critical Phenomena, ed. J. Brey, R.B. Jones, Lecture notes in Physics series, **54**, p.250–273, Springer, Berlin, 1976.

[Ga81] Gallavotti, G.: *Aspetti della Teoria Ergodica Qualitativa e Statistica del Moto*, Quaderno Unione Matematica Italiana, no. **21**, ed. Pitagora, Bologna, 1981, p.1–300.

[Ga89] Gallavotti, G.: *L' hypothèse ergodique et Boltzmann*, in "Dictionnaire Phylosophique", Presses Universitaires de France, p. 1081– 1086, Paris, 1989.

[Ga95a] Gallavotti, G.: *Ergodicity, ensembles, irreversibility in Boltzmann and beyond*, Journal of Statistical Physics, **78**, 1571–1589, 1995.

[Ga95b] Gallavotti, G.: *Trattatello di Meccanica Statistica*, Quaderni CNR-GNFM, **50**, p. 1–350, 1995, Firenze. Free copies can be obtained by requesting them to CNR–GNFM, via S. Marta 13/a, 50139 Firenze, Italy, or by fax at CNR-GNFM -39-55-475915.

[Ga95c] Gallavotti, G.: *Reversible Anosov maps and large deviations*, Mathematical Physics Electronic Journal, MPEJ, (http:// mpej.unige.ch), **1**, 12 pp., 1995.

[Ga95d] Gallavotti, G: *Topics on chaotic dynamics*, in Third Granada Lectures in *Computational Physics*, Ed. P. Garrido, J. Marro, in Lecture Notes in Physics, Springer Verlag, **448**, p. 271–311, 1995.

[Ga95e] Gallavotti, G.: *Meccanica Statistica*, "Quaderni del CNR-GNFM", vol. **50**, p. 1–350, Firenze, 1995.

[Ga96a] Gallavotti, G.: *Chaotic hypothesis: Onsager reciprocity and fluctuation–dissipation theorem*, Journal of Statistical Physics, **84**, 899–926, 1996.

[Ga96b] Gallavotti, G.: *Extension of Onsager's reciprocity to large fields and the chaotic hypothesis*, Physical Review Letters, **78**, 4334–4337, 1996.

[Ga96c] Gallavotti, G.: *Equivalence of dynamical ensembles and Navier Stokes equations*, Physics Letters, **223**, 91–95, 1996.

[Ga97a] Gallavotti, G.: *Dynamical ensembles equivalence in fluid mechanics*, Physica D, **105**, 163–184, 1997.

[Ga97b] Gallavotti, G.: *Chaotic principle: some applications to developed turbulence*, Journal of Statistical Physics, **86**, 907–934, 1997.

[Ga97a] Gallavotti, G., Ruelle, D., : *SRB states and nonequilibrium statistical mechanics close to equilibrium*, Communications in Mathematical Physics, **190**, 279–285, 1997.

[Ga98a] Gallavotti, G.: *A local fluctuation theorem*, Physica A, **263**, 39–50, 1999.

[Ga98b] Gallavotti, G.: *Fluctuation patterns and conditional reversibility in nonequilibrium systems*, Annales de l' Institut H. Poincaré, **70**, 429–443, 1999.

[Ga96d] Gallavotti, G.: *New methods in nonequilibrium gases and fluids*, Open Systems and Information Dynamics, Vol. 6, 101–136, 1999, also in chao-dyn #9610018).

[Ga98d] Gallavotti, G.: *Chaotic dynamics, fluctuations, non-equilibrium ensembles*, Chaos, **8**, 384–392, 1998.

[Ga98e] Gallavotti, G.: *Chaotic Hypothesis and Universal Large Deviations Properties*, chao-dyn 9808004 and Documenta Mathematica, extra volume ICM98, vol. I, 1998.

[Ge93] Georgii, H.O.: *The Equivalence of Ensembles for Classical Systems of Particles*, Journal of Statistical Physics, **80**. 1341–1378, 1995.

[Ge98] Gentile, G: *Large deviation rule for Anosov flows*, Forum Mathematicum, **10**, 89–118, 1998.

[Gi67] Ginibre, J.: *Rigorous lower bound on the compressibility of a classical system*, Physics Letters, **24A**, 223–224, 1967. See also: *Continuity of the pressure as a function of the density for some quantum systems*, Physical Review Letters, **24**, 1473–1475, 1973.

[Gi69a] Ginibre, J.: *Simple proof and generalization of Griffiths second inequality*, Physical Review Letters, **23**, 828–830, 1969. See also: *General formulation of Griffiths inequalities*, Communications in Mathematical Physics, **16**, 310–328, 1970. And *GKS to plane rotators*, Cargese Lecture notes, vol **4**, ed. D. Kastler, Gordon–Breach, Paris, 1970.

[Gi69b] Ginibre, J.: *Existence of phase transitions in quantum lattice systems*, Communications in Mathematical Physics, **14**, 205–234, 1969. See also: *dilute quantum systems*, in *Phase transitions and critical points*, edited by C.Domb and M.S.Green, p. 112–137, Wiley, New York, 1972.

[Gi70] Ginibre, J.: *Systèmes à un nombre infini de degrés de liberté*, Colloques du CNRS, ed. L. Michel, D.Ruelle, Paris, 1970.

[Gi81] Gibbs, J.W.: *Elementary principles in statistical mechanics*, Ox Bow Press, Woodbridge (Connecticut), 1981 (reprint of the 1902 edition).

[Gr58] Grad, H.: *Principles of the kinetic theory of gases*, in *Handbuch der Physik*, vol. XII, p. 205–294, ed. S. Flügge, Springer–Verlag, Heidelberg, 1958.

[Gr62] Groeneveld, J.: Two theorems on classical many particles systems, Physics Letters, **3**, 50–51, 1962.

[Gr67] Griffiths, R.B.: *Correlations in Ising ferromagnets, I*, Journal of Mathematical Physics, **8**, 478–483, 1967.

[Gr70] Griffiths, R.B.: *Concavity of magnetization*, in *A critical review of Thermodynamics*, ed. E.B. Stuart, B. Gal-or, A.J.Brainard, Mono Book Corporations, Baltimore, 1970.

[Gr72] Griffiths, R.B.: *Rigorous results and theorems*, in *Phase transitions and critical points*, edited by C.Domb and M.S.Green, p. 7–109, Wiley, New York, 1972.

[HHP87] Holian, B.L., Hoover, W.G., Posch. H.A.: *Resolution of Loschmidt's paradox: the origin of irreversible behavior in reversible atomistic dynamics*, Physical Review Letters, **59**, 10–13, 1987.

[HKU63] Hemmer, P., Kac, M., Uhlenbeck, G.E. *On the van der Waals theory of the liquid-vapor equilibrium, I,II,III*, Journal of Mathematical Physics, **4**, 216–228, 1963; **4**, 229–247, 1963; **5**, 60–84, 1964.

[HL76] Hemmer, P.C., Lebowitz, J.L.: *Systems with weak long range potentials*, in *Phase transitions and critical points*, vol. **4**, edited by C.Domb and M.S.Green, p. 108–198, Wiley, New York, 1976.

[HS92] Hara, T., Slade, G.: *The lace expansion for self–avoiding walk in five or more dimensions*, Reviews in Mathematical Physics, **4**, 235–327, 1992.

[He95a] Helmholtz, H.: *Principien der Statik monocyklischer Systeme*, in "Wissenschaftliche Abhandlungen", vol. III, p. 142–162 and p. 179– 202, Leipzig, 1895.

[He95b] Helmholtz, H.: *Studien zur Statik monocyklischer Systeme*, in "Wissenschaftliche Abhandlungen", vol. III, p. 163–172 and p. 173– 178, Leipzig, 1895.

[Hi81] Higuchi, Y.: *On the absence of non translationally invariant Gibbs states for the two dimensional Ising system*, in "Random fields", editors J. Fritz, J.L. Lebowitz and D. Szaz, North Holland, Amsterdam, 1981.

[Hi97] Higuchi, Y.: *Ising percolation*, Sugaka Exposition, **10**, 143–158, 1997 (translation of the American Mathematical Society).

[Hu38] Hulthén, L.: Über das Austauschproblem eines Kristalles, Arkiv för Matematik, Astronomi och Fysik, **26A**, 1–106, 1938.

[Hu63] Huang, K.: *Statistical Mechanics*, Wiley, New York, 1963.

[IM65] Ito, K., McKean, H.: *Diffusion processes and their sample paths*, Springer Verlag, New York, 1965.

[Is84] Isakov, S.N.: *Non analytic features of the first order phase transition in the Ising model*, Communications in Mathematical Physics, **95**, 427–443, 1984.

[Ja84] Jacobs, K.: *Ergodic theory and combinatorics*, in Proceedings of the conference on *Modern analysis ansd probability*, june 1982. Contemporary Mathematics, **26**, 171–187, 1984.

[Jo95] Johansson, K.: *On the separation of phase on one dimensional gases*, Communications in Mathematical Physics, **169**, 521–561, 1995.

[KM86] Kotecky, R., Miracle–Solé, S.: *A model with roughening transition at low temperatures*, Physical Review: **34 B**, 2049–2051, 1986.

[KM87] Kotecky, R., Miracle–Solé, S.: *Roughening transitionn for the Ising model on a BCC lattice. A case in the theory of ground states*, Journal of Statistical Physics, **47**, 773–799, 1987. And Kotecky, R., Miracle–Solé, S., Ruiz, J.: *Roughening transition for the Ising model on a BCC lattice*, Journal of Statistical Physics, **47**, 773–799, 1987.

[KO49] Kaufman, B., Onsager, L.: Comment to the "theory of regular solutions" following the paper by G.S. Rushbrooke, Supplemento Nuovo Cimento, **6**, 261, 1949.

[KO93] Kotecky, R., Olivieri, E.: *Droplet dynamics for asymmetric Ising model*, Journal of Statistical Physics, **70**, 1121–1148, 1993.

[KP86] Kotecky, R., Preiss, D.: *cluster expansion for abstract polymers*, Communications in Mathematical Physics, **103**, 491–498, 1986.

[KW41] Kramers H.A., Wannier, G.H.: *Statistics of the two dimensional ferromagnetic Ising model*, Physical Review, **60**, 252–276, 1941.

[Ka49] Kaufman, B.: *Crystal statistics. II. Partition function evaluation by spinor analysis*, Physical Review, **76**, 1232–1243, 1949. See also Kaufman, B., Onsager, L.: *Crystal Statistics. III. Short range order in a binary Ising lattice* Physical Review, **76**, 1244–1252, 1949.

[Ka61] Kasteleyn, P.W.: *The statistics of dimers on a lattice*, Physica, **27**, 1209–1225, 1961.

[Ka68] Kasteleyn, P.W.: *Phase transitions* , in *Fundamental Problems in Statistical Mechanics. II*, edited by E.G.D. Cohen, p. 30–70, North Holland, Amsterdam, 1968.

[Ka69] Kadanoff, L.: *Correlations along a line in the two–dimensional Ising model*, Physical Review, **188**, 859–863, 1969. See also Kadanoff, L., Ceva, H.: *Determination of an operator algebra for the two dimensional Ising model*, Physical Review, **3 B**, 3918–3939, 1971.

[Ka74] Kasteleyn, P.W.: *Combinatorial lattice problems in statistical mechanics*, Proceedings of the *International school of Mathematical Physics*, p. 101–136, ed. G. Gallavotti, Camerino, 1974, University of Camerino Press, 1976.

[Kl62] Klein, M.: *Maxwell and the beginning of the Quantum Theory*, Archive for the history of exact sciences, **1**, 459–479, 1962.

[Kl72] Klein, M.: *Mechanical explanations at the end of the nineteenth century*, Centaurus, **17**, 58–82, 1972.

[Kl73] Klein, M.: *The development of Boltzmann statistical ideas*, in "The Boltzmann equation", ed. E.G.D Cohen, W. Thirring, Acta Physica Austriaca, suppl. X, Wien, p. 53–106, 1973.

[Kr79] Krylov, N.S.: *Works on the foundations in statistical physics*, Princeton University Press, 1979, (english translations of selected papers).

[Ku83] Kurota, K.: *Phase separations in Ising model with free boundary conditions*, Journal of Statistical Physics, **36**, 1–13, 1983.

[Ku87] Kuhn, T.: *Black body theory and the quantum discontinuity. 1814–1912*, University of Chicago Press, 1987.

[Ku97] Kurchan, J.: *Fluctuation Theorem for stochastic dynamics*, Journal of Physics, A, **31**, 3719–3729, 1998.

[LL63], Lieb, E., Lininger, W.: *Exact analysis of an interacting Bose gas. I. The general solution and the ground state*, Physical Review, **130**, 1605–1616, 1963; reprinted in [LM66].

[LL67] Landau, L., Lifschitz, L.E.: *Physique Statistique*, MIR, Moscow, 1967.

[LL72] Lieb, E.H., Lebowitz, J.L.: *Lectures on the therodynamic limit for Coulomb systems*, in *Springer lecture notes in Physics*, edited by A. Lenard, vol. **20**, p. 135–161, Berlin, 1972.

[LLP97] Lepri. S., Livi, R., Politi, A. *Energy transport in anharmonic lattices close and far from equilibrium*, preprint, cond-mat@xyz. lanl. gov #9709156, to appear on Physical Review E.

[LM66] Lieb, E., Mattis, D.C.: *Mathematical Physics in one dimension*, Academic Press, New York, 1966.

[LMMRS91] Laanait, L., Messager, A., Miracle–Solé, S., Ruiz, J., Shlosman, S.: *Interfaces in the Potts model. (I) Pirogov-Sinai theory of the Fortuin-Kasteleyn representation.* Communications of Mathematical Physics, **140**, 81–91, 1991.

[LMP98] Lebowitz, J.L., Mazel, A., Presutti, E.: *Rigorous proof of a liquid-vapor phase transition in a continuum particle system*, in print on Physical Review Letters, 1998.

[LP66] Lebowitz, J., Percus, J.: *Kinetic equations and exactly solvable one dimensional systems*, Physical Review, **155**, 122–138, 1966; see also J. Lebowitz, J. Percus, J. Sykes, *Time evolution of the total distribution function of a one dimensional systems of hard rods*, Physical Review, **171**, 224–235, 1968.

[LP76] Lebowitz, J.L., Presutti, E.: *Statistical Mechanics of systems of unbounded spins*, **50**, 195–218, 1976.

[LP79] Lebowitz, J.L., Penrose, O.: *Towards a rigorous molecular theory of metastability*, in *Fluctuation phenomena*, ed. E.W. Montroll, J.L. Lebowitz, North Holland, Amsterdam, 1979.

[LR69] Lanford, O., Ruelle, D.: *Observables at infinity and states with short range correlations in statistical mechanics*, Communications in Mathematical Physics, **13**, 194–215, 1969.

[LS82] Lebowitz, J., Spohn, H.: *Steady state self diffusion at low density*, Journal of Statistical Physics, **29**, 39–55, 1982.

[LS94] Liddell, H., Scott, R.: *Greek–English lexicon*, Oxford, 1994.

[LS98] Lebowitz, J.L., Spohn, H.: *The Gallavotti-Cohen Fluctuation Theorem for Stochastic Dynamics*, Journal of Statistical Physics, **95**, 333–365, 1999.

[LT75] Lieb, E., Thirring, W.: *Bound for the kinetic energy of fermions which proves the stability of matter*, Physical Review Letters, **11**, 687–689, 1975.

[LV93] Levesque, D., Verlet, L.: *Molecular dynamics and time reversibility*, Journal of Statistical Physics, **72**, 519–537, 1993.

[LW72] Lieb, E., F.Y. Wu: *Two Dimensional Ferroelectric Models*, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. Green Vol. **1**, p. 332-490, Academic Press, London, 1972.

[LY52] Lee T.D., Yang C.N.: *Statistical theory of equations of state and phase transitions, II. Lattice gas and Ising model*, Physical Review, **87**, 410–419, 1952.

[LY73] Lasota, A., Yorke, J.A.: *On the existence ungheria.tex-of invariant measures for piecewise monotonic transformations*, ungheria.tex-Transactions American Mathematical Society, **183**, 481, 1973.

[La68] Lanford, O.: *The classical mechanics of one dimensional systems of infinitely many particles.I. An existence theorem*, Communications in Mathematical Physics, **9**, 176–191, 1968.

[La69] Lanford, O.: *The classical mechanics of one dimensional systems of infinitely many particles.II. Kinetic theory*, Communications in Mathematical Physics, **11**, 257–292, 1969.

[La70], Lanford, O.: *Quantum Spin Systems*, in "Cargese Lectures in Physics", vol. IV, ed. D. Kastler, , pp. 113–115, Gordon Breach, Paris, 1970.

[La72], Lanford, O.: *Entropy and equilibrium states in classical statistical mechanics*, in *Springer lecture notes in Physics*, edited by A. Lenard, vol. **20**, p. 1–113, Berlin, 1972.

[La74] Lanford, O.: *Time evolution of large classical systems*, in "Dynamical systems, theory and applications", p. 1–111, ed. J. Moser, Lecture Notes in Physics, vol. 38, Springer Verlag, Berlin, 1974.

[La85] Lawler, G.F.: *Intersections of random walks in four dimensions*, Communications in Mathematical Physics, **97**, 583–594, 1985. See also *Intersections of simple random walks* in "Particle systems, random media and large deviations", Series 'Contemporary Mathematics', vol. **41**, American Mathematical Society, Providence, 1985.

[Le74] Lebowitz, J.L.: *GHS and other inequalities*, Communications in Mathematical Physics, **28**, 313–321, 1974.

[Le93] Lebowitz, J.L.: *Boltzmann's entropy and time's arrow*, Physics Today, 32–38, 1993. And *Microscopic Reversibility and Macroscopic Behavior: Physical Explanations and Mathematical Derivations*, in 25 years of non-equilibrium statistical mechanics, ed. J. Brey, J. Marro, J. Rubi, M. San Miguel, Lecture Notes in Physics, **445**, Springer, Berlin, 1995.

[Li67a] Lieb, E.: *Exact Solution of the Problem of the Entropy of Two-Dimensional Ice*: Physical Review Letters, **18**, 692–694, 1967. And *The Residual Entropy of Square Ice*, Physical Review, **162**, 162–172, 1967.

[Li67b] Lieb, E.: *Exact Solution of the F Model of an Antiferroelectric*, Physical Review Letters, **18**, 1046–1048, 1967;

[Li67c] Lieb, E.: *Exact Solution of the Two-Dimensional Slater KDP Model of a Ferroelectric*, Physical Review Letters, **19**, 108–110, 1967.

[Li81] Lieb, E: *Thomas–Fermi and related theories of atoms and molecules*, Reviews of Modern Physics, **53**, 603–641, 1981.

[Li96] Liverani, C.: *Deacy of correlations*, Annals of Mathematics, **142**, 239–301, 1995.

[ML72] Martin–Löf, A., Lebowitz, J.L.: *On the uniqueness of the equilibrium state for Ising spin systems*, Communications in Mathematical Physics, **25**, 276–282, 1972.

[MM77] Messager, A., Miracle–Solé, S.: *Correlation functions and boundary conditions in the Ising ferromagnet*, Journal of Statistical Physics, **17**, 245–262, 1977.

[MMR92] Messager, A., Miracle–Solé, S., Ruiz, J.: *Convexity properties of surface tension and equilibrium crystals*, Journal of Statistical Physics, **67**, 449–470, 1992.

[MMRS91] Messager, A., Miracle–Solé, S., Ruiz, J., Shlosman, S.: *Interfaces in the Potts model. (II) Antonov rule and rigidity of the order-disorder interface.* Communications in Mathematical Physics **140**, 275–290, 1991.

[MOS90] Martinelli, F., Olivieri, E, Scoppola, E.: *Metastability and exponential approach to equilibrium for low temperature stochastic Ising models*, J. Stat. Physics, **61**, 1105–1119, 1990.

[MP72] Marchioro, C., Presutti, E.: *Thermodynamics of particle systems in presence of external macroscopic fields*, Communications in Mathematical Physics, **27**, 146–154, 1972. And **29**, 265–284, 1973.

[MR94] Miracle–Solé, S., Ruiz, J.: *On the Wulff construction as a problem of equivalence of ensembles*, in Micro, Meso and Macroscopic Approaches in Physics, ed. M. Fannes, A. Verbeure, Plenum, New York, 1994.

[MR96] Morriss, G.P., Rondoni, L.: *Equivalence of "nonequilibrium" ensembles for simple maps*, Physica A, **233**, 767-784, 1996.

[MR97] Morriss, G.P., Rondoni, L.: *Applications of periodic orbit theory to N–particle systems*, Journal of Statistical Physics, **86**, 991–1009, 1997.

[MS67] Minlos, R.A., Sinai, J.G.: *The phenomenon of phase separation at low temperatures in some lattice models of a gas, I*, Math. USSR Sbornik, **2**, 335–395, 1967. And: *The phenomenon of phase separation at low temperatures in some lattice models of a gas, II*, Transactions of the Moscow Mathematical Society, **19**, 121–196, 1968. Both reprinted in [Si91].

[MS70] Minlos, R.A., Sinai, J.G.: *Investigation of the spectra of stochastic operators arising in lattice gas models*, (russian) Teoreticheskaia y Matematicheskaia Fyzika, **2**, No. 2, 230–243, 1970.

[MT77] McCoy, B.M., Tracy, C., Wu, T.T.: *Two dimensional Ising model as an explicitly solvable relativistic field theory: explict formulas for n–point functions*, Physical Review Letters, **18**, 1253–1258, 1978.

[MW66] Mermin, N.D., Wagner, H. *Absence of ferromagnetism or antiferromagnetism in* 1 *or* 2 *diensional isotropic Heisenberg models*, Physical Review Letters, **17**, 1183–1186, 1966.

[MW73] McCoy, B.M., Wu, T.T.: *The two dimensional Ising model*, Harvard University Press, Cambridge, 1973.

[MW86] McCoy, B.M., Wu, T.T.: *Some exact results for the electric correlation functions for the eight vertex model*, Physical Review Letters, **56**, 2645–2648, 1986.

[Ma65] Mattis, D.C.: *Theory of Magnetism*, Harper–Row, New York, 1965.

[Ma72] Martin–Löf, A: *On the spontaneous magnetization in the Ising model*, Communications in Mathematical Physics, **24**, 253–259 ,1971.

[Ma79] Maxwell, J.: *On Boltzmann's theorem on the average distribution of energy in a system of material points*, in "The scientific papers of J,C, Maxwell", ed. W. Niven, Cambridge University Press, 1890, vol. II, p. 713–741.

[Ma88] Mathieu, M.: *On the origin of the notion 'Ergodic Theory'*, Expositiones Mathematicae, **6**, 373–377, 1988.

[Ma98] Mastropietro, V.: *Renormalization group and correlation functions for the XYZ model*, preprint, Roma, 1998.

[Me67] Mermin, D.: *Absence of ordering in certain classical systems*, Journal of Mathematical Physics, **8**, 1061–1064, 1967.

[Me71] Mermin, D.: *Solvable model of a vapor–liquid transition with sigular coexistence–curve diameter*, Physical Review Letters, **26**, 168–172, (1971).

[Mi68] Minlos, R.A.: *Lectures in Statistical Mechanics*, in *Russian Mathematical Surveys*, **23**, 137–196, 1968.

[Mi95] Miracle–Solé, S.: *On the microscopic theory of phase coexistence*, *25 Years of Non-Equilibrium Statistical Mechanics*, ed. J.J. Brey, Lecture Notes in Physics, **445**, pp. 312–322. Springer, Berlin, 1995. And *Surface tension, step free energy and facets in the equilibrium crystal shape*, Journal Statistical Physics, **79**, 183–214, 1995.

[Mo56] Morrey, C.B.: *On the derivation of the equations of hydrodynamics from Statistical Mechancs*, Communications in Pure and Applied Mathematics, **8**, 279–326, 1955.

[NM53] G.F. Newell G.F., Montroll E.W.: *On the theory of the Ising model of ferromagnetism*, Reviews of Modern Physics, **25**, 353–389, 1953.

[Ne67] Nelson, E.: *Dynamical theories of Brownian motion*, Princeton University Press, 1967.

[On44] Onsager L.: *Crystal statistics. I. A two dimensional Ising model with an order–disorder transition*, Physical Review, **65**, 117–149, 1944.

[Or58] Orbach, R.: *Linear antiferromagnetic chain with anisotropic coupling*, Physical Review, **112**, 309–316, 1958.

[Or74] Ornstein, D.: *Ergodic theory, randomness and dynamical systems*, Yale University Press, 1974.

[PS76] Pirogov, S.A., Sinai, J.G.: *Phase diagrams of classical lattice systems*, Theoretical and Mathematical Physics, **25**, 1185–1192, 1975; and **26**, 39–49, 1976. Both reprinted in [Si91].

[PV96] Pfister, C. Velenik, Y.: *Mathematical theory of the wetting phenomenon in the* 2*D–Ising model*, Helvetica Physica Acta, **69**, 949–973, 1996; Pfister, C., Velenik, Y.: *Large deviations and continuum limit in the 2D Ising model*, Probability Theory and Related fields, **109**, 435–506, 1997.

[PV99] Pfister, C. Velenik, Y.: *Interface, surface tension and reentrant pinning transition in the* 2*D Ising model*, Communications in Mathematical Physics, **204**, 269–312, 1999.

[Pa82] Pais, A.: *Subltle is the Lord: the science and the life of Albert Einstein*, Oxford University Press, 1982.

[Pe36] Peierls, R.:*On Ising's model of ferromagnetism*, Proceedings of the Cambridge Philosophical Society, **32**, 477–481, 1936.

[Pe63] Penrose, O.: *Convergence of the fugacity expansions for fluids and lattice gases*, Journal of Mathematical Physics, **4**, 1312–1320, 1963. See also *The remainder in Mayer's fugacity series*, Journal of Mathematical Physics, **4**, 1488–1494, 1963.

[Pe70] Perrin, J.: *Les atomes*, Gallimard, Paris, 1970 (reprint).

[Po00] Poincaré, H.: *Relations entre la Physique expérimentale et la Physique mathématique*, in *La Science et l'Hypothèse*", p. 157–190, Champs–Flammarion, Paris, 1995. And: *The crises of mathematical physics*, in *La valeur de la Science*, p. 129–150, Champs–Flammarion, Paris, 1994.

[RM75] Russo, L., Monroy, G.: *A family of codes between some Markov and Bernoulli schemes*, Communications in Mathematical Physics, **43**, 153–160, 1975.

[RT71] Ruelle, D., Takens, F.: *On the nature of turbulence*, **20**, 167–192, 1971. And *Note concerning our paper "On the nature of turbulence"*, Communications in Mathematical Physics, **23**, 343–344, 1971: this note is important as it brings to attention several references that were not well known at the time.

[RT98] Rom–Kedar, V., Turaev, D.: *Big islands in dispersing billiard–like potentials*, preprint, Weizmann Institute, april 2, 1998.

[Ru64] Ruelle, D.: *Cluster property of the correlation functions of classical gases*, Reviews of Modern Physics, **36**, 580–584, 1964.

[Ru68] Ruelle, D.: *Statistical mechanics of one–dimensional lattice gas*, Communications in Mathematical Physics, **9**, 267–278, 1968.

[Ru69] Ruelle, D.: *Statistical Mechanics*, Benjamin, New York, 1969.

[Ru70] Ruelle, D.: *Superstable interactions in classical statistical mechanics*, Communications in Mathematical Physics, **18**, 127–159, 1970. See also: *Probability estimates for continuous spin systems*, Communications in Mathematical Physics, **50**, 189–194, 1976.

[Ru71a] Ruelle, D.: *Extension of the Lee–Yang circle theorem*, Physical Review Letters, **26**, 303–304, 1971.

[Ru71b] Ruelle, D.: *Existence of a phase transition in a continuous system*, Physical Review Letters: **27**, 1040–1041 (1971).

[Ru73b] Ruelle, D.: *On the use of small external fields in problems of symmetry breakdown in statistical mechanics*, Annals of Physics, **69**, 364–374, 1972.

[Ru76] Ruelle, D.: *A measure associated with Axiom A attractors*, American Journal of Mathematics, **98**, 619–654, 1976.

[Ru78a] Ruelle, D.: *Sensitive dependence on initial conditions and turbulent behavior of dynamical systems*, Annals of the New York Academy of Sciences, **356**, 408–416, 1978.

[Ru78b] Ruelle, D.: *Thermodynamic formalism*, Addison Wesley, Reading, 1978.

[Ru78c] Ruelle, D.: *What are the measures describing turbulence?*, Progress of Theoretical Physics, (Supplement) **64**, 339–345, 1978.

[Ru79a] Ruelle, D.: *Ergodic theory of differentiable dynamical systems*, Publications Mathématiques de l' I.H.E.S, **50**, p. 275–320, 1979.

[Ru79b] Russo, L.: *The infinite cluster method in the two dimensional Ising model*, Communications in Mathematical Physics, **67**, 251–266, 1979.

[Ru80] Ruelle, D.: *Measures describing a turbulent flow*, Annals of the New York Academy of Sciences, volume *Nonlinear dynamics*, ed. R.H.G. Helleman, **357**, 1–9, 1980.

[Ru81] Russo, L.: *On the critical percolation probabilities*, Zeitschrift für Warsheinslickeit und werwandter Gebiete, **56**, 229–237, 1981.

[Ru89] Ruelle, D.: *Elements of differentiable dynamics and bifurcation theory*, Academic Press, London, 1989.

[Ru96a] Ruelle, D.: *Positivity of entropy production in nonequilibrium statistical mechanics*, Journal of Statistical Physics, **85**, 1–25, 1996.

[Ru97a] Ruelle, D.: *Entropy production in nonequilibrium statistical mechanics*, Communications in Mathematical Physics, **189**, 365–371, 1997.

[Ru97b] Ruelle, D.: *Differentiation of SRB states*, Communications in Mathematical Physics, **187**, 227–241, 1997. And *Nonequilibrium statistical mechanics near equilibrium: computing higher order terms*, preprint, IHES, july 1997.

[Ru97c] Ruelle, D.: *Smooth dynamics and new theoretical ideas in non-equilibrium statistical mechanics*, Journal of Statistical Physics, **95**, 93–468, 1999.

[SJ93] She, Z.S., Jackson, E.: *Constrained Euler system for Navier–Stokes turbulence*, Physical Review Letters, **70**, 1255–1258, 1993.

[SML64] Schultz T.D. , Mattis, D.C., Lieb, E.: *The two dimensional Ising model as a soluble problem of many fermions*, Reviews of Modern Physics, **36**, 856–871, 1964.

[Sc86] Schwartz, J.: *The Pernicious Influence of Mathematics on Science*, in "Discrete thoughts: essays in Mathematics, Science, and Phylosophy", M. Kac, G. Rota, and J. Schwartz, eds., Birkhauser, Boston, 1986, p. 19–25.

[Si68] Sinai, Y.G.: *Markov partitions and C-diffeomorphisms*, Functional Analysis and Applications, **2**, 64–89, 1968, n.1; and: *Construction of Markov partitions*, Functional analysis and Applications, **2**, 70–80, 1968, n.2.

[Si70] Sinai, Y.: *Dynamical systems with elastic reflections. Ergodic properties of dispersing billards*, Russian Mathematical Surveys, **25**, 137–189, 1970.

[Si72] Sinai, Y.G.: *Gibbs measures in ergodic theory*, Russian Mathematical Surveys, **27**, 21–69, 1972.

[Si77] Sinai, Y.G.: *Lectures in ergodic theory*, Lecture notes in Mathematics, Princeton U. Press, Princeton, 1977.

[Si79] Sinai, Y.G.: *Development of Krylov's ideas*, in Krylov, N.S.: *Works on the foundations in statistical physics*, Princeton University Press, 1979, p. 239–281.

[Si91] Sinai Ya.G.: *Mathematical problems of statistical mechanics*, World Scientific, 1991.

[Sp71] Spitzer, F.: *Markov random fields and Gibbs ensembles*, American Mathematical Monthly, **78**, 142–154, 1971.

[Sp91] Spohn, H.: *Large scale dynamics of interacting particles*, Springer Verlag, Berlin, 1991.

[Su67] Sutherland, W.: *Exact solution of a two dimensional model for hydrogen–bonded crystals*, Physical Review Letters, **19**, 103–104, 1967.

[Su70] Sutherland, W.: *Two-Dimensional Hydrogen Bonded Crystals without the Ice Rule*, Journal of Mathematical Physics, **11**, 3183–3186, 1970.

[KSiS95] Krámli, A., Simányi, N., Szász, D.: *Ergodic properties of semi–dispersing billiards*, Nonlinearity, **2**, 311–326, 1995. See also Szász, D.: *Boltzmann's ergodic hypothesis, a conjecture for centuries?*, Studia Sci. Math. Hungaricae, **31**, 299–322, 1996.

[TF61] Temperley, H.N.V., Fisher, M.E.: *Dimer problem in Statistical Mechanics. An exact result*, Philosophical Magazine, **6**, 1061–1063, 1961.

[Ta66] Taylor, W.: *The kinetic theory of the dissipation of energy*, reprinted (in english) in: S. Brush, *Kinetic theory*, **2**, 176, Pergamon Press, Oxford, 1966.

[Th74] Thomson, W.: *The kinetic theory of dissipation of energy*, Proceedings of the Royal Society of Edinburgh, **8**, 325-328, 1874, reprinted in [Br66].

[UF63] Uhlenbeck, G.E., Ford, G.W.: *Lectures in Statistical Mechanics*, American Mathematical society, Providence, R.I., 1963, pp. 5,16,30.

[UO30] Uhlenbeck, G.E., Ornstein, L.S.: *On the theory of Brownian motion*, Physical Review, **36**, 823–841, 1930, reprinted in [Wa54].

[Ul68], Uhlenbeck, G.E.: *An outline of Statistical Mechanics*, in *Fundamental problems in Statistical Mechanics*, vol. II, ed. E.G.D. Cohen, North Holland, Amsterdam, 1968.

[VB75] van Beijeren, H.: *Interface sharpness in the Ising system*, Communications in Mathematical Physics, **40**, 1–6, 1975.

[VB77] van Beijeren, H.: *Exactly solvable model for the roughening transition of a crystal surface*, Physical Review Letters, **38**, 993–996, 1977.

[VH50] van Hove, L.: *Sur l'intégrale de configuration pour les systèmes de particules à une dimension*, Physica, **16**, 137–143, 1950.

[VK64] van Kampen, N.G.: *COndensation of a classical gas with long range attraction*, Physical Review, **135A**, 362–369, 1964.

[VN55] von Neumann, J.: *Mathematical foundations of quantum mechanics*, Princeton University Press, Princeton, New Jersey, 1955.

[VN79] von Nägeli, K.W.: *Über die Bewegungen kleinster Körper*, Bayerische Akademie der Wissenschaften, Sitzungsberichte der Mathematisch-Physikalischen Classe, **9**, 389–453, 1879.

[VN87] van Beijeren, H., Nolden, I.M.: *The roughening transition*, in *Topics in Current Physics*, ed. W. Schommers, P. von Blankenhagen, Springer, 1987.

[VP92] von Plato, J.: *Boltzmann's ergodic hypothesis*, Archive for the History of exact sciences, **44**, 71-89, 1992.

[VP94] von Plato, J.: *Creating modern probability*, Cambridge University Press, New York, 1994.

[VS06] von Smoluchowski, M.: *Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen*, Annalen der Physik, **21**, 756–780, 1906.

[VW41] van der Waerden, B.L.: *Die lange reichweite der regelmässigen atomanordnung in Mischcristallen*, Zeitschrift für Physik, **118**, 473–488, 1941.

[VW88] van der Waals, J.D.: *On the continuity of the gaseous and liquid states*, thesis, 1873, edited by J.S. Rowlinson, in *Studiesin Statistical Mechanics*, vol. XIV, North Holland, Amsterdam, 1988.

[WF72] Wilson, K.G., Fisher, M.E.: *Critical exponents in* 3.99 *dimensions*, Physical Review Letters, **28**, 240–243, 1972.

[WL69] A. Weijland, J. Van Leeuwen, *Non analytic behaviour of the diffusion coefficient of a Lorentz gas*, Physica, **36**, 457–490, 1967; and **38**, 35–47, 1968; E. Hauge, E. Cohen, *Normal and abnormal diffusion in Ehrenfests's wind tree model*, Journal of Mathematical Physics, **10**, 397–414, 1969; W., Hogey, *Convergent generalizations of the Boltzmann equation for a hard sphere gas*, Physical Review, **185**, 210–218, 1969.

[WL98] Woitkowsky, M.P., Liverani, C.: *Conformally Symplectic Dynamics and Symmetry of the Lyapunov Spectrum*, Communications in Mathematical Physics, **194**, 47–60, 1998.

[WMTB76] Wu, T.T., McCoy, B.M., Tracy, C., Baruch, E.: *Spin–spin correlation functions for the two-dimensional Ising model: exact theory in the scaling region*, Physical Review **13 B**, 316–374, 1976.

[Wa54] Wax, N.: *Selected papers on noise and stochastic processes*, Dover, 1954.

[Wa59] Walker, L.R.: *Antiferromagnetic linear chain*, Physical Review, **116**, 1089–1090, 1959.

[We71] Wegner, F.J.: *Duality in generalized Ising models and phase transitions without local order parameters*, Journal of Mathemathical Physics, **12**, 2259–2272, 1971.

[Wi83] Wilson, K.G.: *The renormalization group and critical phenomena*, Review of Modern Physics, **55**, 583–600, 1983.

[Wi89] Wittaker, E.T.: *A treatise on the analytic dynamics of particles & rigid bodies*, Cambridge University Pres, 1989, (reprint).

[YY66] Yang, C.N., Yang, C.P.: *Ground state energy of a Heisenberg–Ising lattice*, Physical Review, **147**, 303–306, 1966; Physical Review, **150**, 321–327, 327–339, 1966; Physical Review, **151**, 258–264, 1966.

[Ya52] Yang, C.N.: *The spontaneous magnetization of a two–dimensional Ising model*, Physical Review, **85**, 808–816, 1952.

[Yo94] Young, L.S.: *Ergodic theory of differentiable dynamical systems*, in *Real and complex dynamical systems*, ed. B. Branner, P. Hjorth, Nato ASI series, Kluwer, Dordrecht, 1995. And *Ergodic theory of chaotic dynamical systems*, Proceedings of the International Congress of Mathematical Physics, Brisbane, 1997, ed. A. Bracken.

[Za89] Zanetti, G.: *Hydrodynamics of lattice gas automata*, Physical Review, **40 A**, 1539–1548, 1989.

Names index

A

B

C

D

E

F

G

H

J

K

L

M

N

O

P

Analytic index

V

w

Z

Citations index